

Cub - 110294-14 - 0335 088

AMERICAN PHILOSOPHICAL QUARTERLY

14

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George A. Schreier
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 14/NUMBER 1

JANUARY, 1977

CONTENTS

- | | | | |
|--|----|---|----|
| I. PETER ACHINSTEIN: <i>What is an Explanation?</i> | I | V. MERRILL RING: <i>Knowledge The Cessation of Belief</i> | 51 |
| II. ALAN H. GOLDMAN: <i>Justice and Hiring by Competence</i> | 17 | VI. JOHN D. HODSON: <i>The Principle of Paternalism</i> | 61 |
| III. CAROLYN R. MORILLO: <i>Doing, Refraining, and the Strenuousness of Morality</i> | 29 | VII. DAN W. BROCK: <i>The Justification of Morality</i> | 71 |
| IV. THOMAS D. PERRY: <i>A Paradigm of Philosophy: Hohfeld on Legal Rights</i> | 41 | VIII. RAYMOND DACEY, RICHARD E. SIMMONS, DAVID J. CURRY, JOHN W. KENNELLY: <i>A Cognitive Solution to Newcomb's Problem</i> | 79 |

© AMERICAN PHILOSOPHICAL QUARTERLY, 1977

PUBLISHED IN OXFORD BY BASIL BLACKWELL

ISSN 0003-0481

American

PHILOSOPHICAL QUARTERLY

105

Am 39

POLICY

American Philosophical Quarterly welcomes articles by philosophers of any country on any aspect substantive or historical. However, only self-sufficient articles will be published, and not book reviews, critical notices, or "discussion notes" (short or long). Moreover the editor will consider papers whose authors certify that while under consideration with us, they will not be submitted elsewhere.

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The subscription for 1977 is \$28.00 for institutions, \$12.00 to individuals. Single issues will be \$8.00 to all purchasers. For Monograph prices, see below. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The complete series is listed on the back cover. In 1977 these will be available at \$9.00 to institutions and non-subscribers, \$4.50 to individual subscribers. Please note that it is no longer possible to supply the monographs free of charge to individual subscribers.

335088



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton



VOLUME 14 (1977)

PUBLISHED IN OXFORD BY BASIL BLACKWELL

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 14 (1977)

ACHINSTEIN, PETER: <i>What Is An Explanation?</i>	I
ADAMS, ROBERT MERRIHEW: <i>Middle Knowledge and the Problem of Evil</i>	109
BRAND, MYLES: <i>Identity Conditions for Events</i>	329
BROCK, DAN W.: <i>The Justification of Morality</i>	71
CHISHOLM, RODERICK M.: <i>Thought and Its Reference</i>	167
COOPER, JOHN M.: <i>The Psychology of Justice in Plato</i>	151
CORNMAN, JAMES W.: <i>Foundational Versus Nonfoundational Theories of Empirical Justification</i>	287
DACEY, RAYMOND <i>et al.</i> : <i>A Cognitivist Solution to Newcomb's Problem</i>	79
DAVIS, LAWRENCE H.: <i>Prisoners, Paradox, and Rationality</i>	319
DAY, J. P.: <i>Threats, Offers, Law, Opinion, and Liberty</i>	257
DUFF, ANTONY: <i>Psychopathy and Moral Understanding</i>	189
GOLDMAN, ALAN H.: <i>Justice and Hiring By Competence</i>	17
GRIMM, ROBERT: <i>Eventual Change and Action Identity</i>	221
HARMAN, GILBERT: <i>How to Use Propositions</i>	173
HARTSHORNE, CHARLES: <i>The Neglect of Relative Predicates in Modern Philosophy</i>	309
HODSON, JOHN D.: <i>The Principle of Paternalism</i>	61
HOLMAN, EMMETT L.: <i>Sensory Experience, Perceptual Evidence, and Conceptual Frameworks</i>	99
JAEGER, ROBERT A.: <i>Am I in the World?</i>	239
KYBURG, HENRY E.: <i>All Acceptable Generalizations Are Analytic</i>	201
MARGOLIS, JOSEPH: <i>Skepticism, Foundationalism, and Pragmatism</i>	119
MCDERMOTT, MICHAEL: <i>Sets As Open Sentences</i>	247
MORILLO, CAROLYN R.: <i>Doing, Refraining, and the Strenuousness of Morality</i>	29
PERRY, THOMAS D.: <i>A Paradigm of Philosophy: Hohfeld on Legal Rights</i>	41
RAWLS, JOHN: <i>The Basic Structure as Subject</i>	159
RING, MERRILL: <i>Knowledge: The Cessation of Belief</i>	51
RUNZO, JOSEPH: <i>The Propositional Structure of Perception</i>	211
RUSSELL, BRUCE: <i>On the Relative Strictness of Negative and Positive Duties</i>	87
SCHLAGEL, RICHARD H.: <i>The Mind-Brain Identity Impasse</i>	231
SIKORA, R. I.: <i>The Argument From Analogy Is Not An Argument For Other Minds</i>	137
SNARE, FRANK: <i>Three Sceptical Theses in Ethics</i>	129
UNGER, PETER: <i>The Uniqueness of Causation</i>	177
VAN FRAASSEN, BAS C.: <i>The Pragmatics of Explanation</i>	143
VAN INWAGEN, PETER: <i>Creatures of Fiction</i>	299
WRIGHT, EDMOND: <i>Perception: A New Theory</i>	273

I. WHAT IS AN EXPLANATION?

PETER ACHINSTEIN

1. Introduction

FOR over a quarter century the philosophical literature on explanation has been dominated by one theory: the deductive-nomological (D-N) model. In the course of formulating conditions an explanation must satisfy to be correct, adherents of this model express a view about the ontological character of explanations, viz. that they are arguments. But ontological questions are not of paramount importance to these theorists or to their critics, who object mainly to conditions such as deducibility and the inclusion of laws which D-N theorists impose on explanations. More generally, in discussions of this as well as of other theories of explanation there has been no airing of ontological views, no general consideration of reasons for or against construing explanations as arguments, propositions, sentences, or perhaps other sorts of entities. It is my aim in this paper to examine these ontological possibilities. I want to show why the simpler candidates will not do, and that if explanations are entities at all they are more complex than is generally supposed. Such an exploration will provide one kind of answer to the question "What is an explanation?" It will also be of interest to discover how dependent on its ontology is a theory that provides conditions for evaluating explanations. Would the D-N model, e.g., be precluded by the adoption of an ontology different from that suggested by the model itself?

Explanations come in all sizes and shapes. One might explain Bill's stomach ache by saying that it is due to something he ate, or one might offer a more elaborate account indicating what he ate and how it caused his stomach ache. In both cases explanations are produced, no matter which we may consider the more adequate. Although the examples used in this paper are almost always quite simple, the theories to be studied are meant to apply to explanations generally, whether simple or

complex, adequate or inadequate. Having developed a suitable ontological view it is always possible to go on to consider how to divide the good from the bad (the scientific from the unscientific, etc.).

I shall begin with a distinction essential for our discussion. Suppose that Dr. Smith explained Bill's stomach ache by saying that Bill ate spoiled meat. There is some *act* that has occurred, viz. the doctor's explaining, which went on for some period of time however short. There is also the *product* of this act, viz. the explanation given by the doctor, which did not go on for any period of time but was produced in or by the act of explaining.¹ My main concern is with the product of explanation, and to discover what this is I shall consider sentences of the form

(1) The explanation of *q* given by *S* is that *p*,

in which "the explanation of *q* given by *S*" will be called the product-expression ("*q*" will be called the object-expression). In (1) '*S*' can be replaced by expressions referring to speakers and '*p*' by sentences. My question is whether the product-expression in sentences of form (1) should be construed as referring, and if so to what.² Before we can turn to product-theories, however, some remarks will need to be made about the explaining act.

2. The Act of Explaining

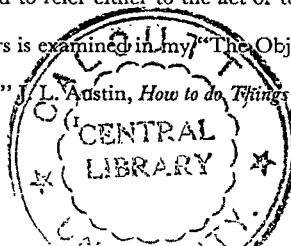
Explaining is an illocutionary act. Like warning and promising it is typically performed by uttering (or writing) words in certain contexts with certain intentions.³ It is to be distinguished from such perlocutionary acts as enlightening someone or getting him to understand, which are effects one's act of explaining can have upon the thoughts and beliefs of others.

The illocutionary character of explaining can be exposed by formulating a set of conditions for per-

¹ "The explanation given by the doctor" can be used to refer either to the act or to the product. In what follows it will be used in the latter way.

² The question of whether the object-expression refers is examined in my "The Object of Explanation," in Stephan Körner (ed.), *Explanation* (Oxford, 1975), pp. 1-45.

³ Austin includes "explain" on his list of "expositives." J. L. Austin, *How to do Things with Words* (Oxford, 1962), pp. 160-161.



forming such an act. To do so I shall consider sentences of the form "*S* explains *q* by uttering *u*" in which '*S*' denotes some person. First, *S* explains *q* by uttering *u* only if

- (1) *S* utters *u* with the intention that his utterance of *u* render *q* understandable.

This condition expresses the point of *S*'s act. I will assume here without argument that there is such a relationship between explaining and understanding.⁴ To explain *q*, however, is not to utter just anything with the intention of rendering *q* understandable. Suppose I believe that the words "truth is beauty" are so causally efficacious with you that the mere uttering of them will cause you to understand anything, including why Bill got a stomach ache. By uttering these words I have not thereby explained why Bill got a stomach ache, even if I have satisfied (1). The reason is that I do not believe that "truth is beauty" expresses a correct answer to the question "Why did Bill get a stomach ache?" More generally, *S* explains *q* by uttering *u* only if

- (2) *S* believes that *u* expresses a correct answer to *Q*. (*Q* is a question which "presupposes" *q*).⁵

These conditions are not yet sufficient, however. Suppose that *S* intends that his uttering of *u* will render *q* understandable by causing people to think of *u'* which, like *u*, *S* believes expresses a correct answer to *Q*. In such a case although *S* utters something which he believes will cause others to be able to explain *q*, *S* does not himself explain *q* by uttering *u*. To preclude this possibility we might say that *S* explains *q* by uttering *u* only if *S* utters *u* with the intention that his utterance of *u* render *q* understandable by producing the recognition that *u* expresses a correct answer to *Q*.

However, a difficulty similar to the one we are trying to avoid will arise even with this condition. We might imagine a case (admittedly bizarre) in which *S* intends his utterance of *u* to render *q* understandable in this indirect way: When others hear the utterance of *u* and recognize that *u* expresses a correct answer to *Q*, they will be caused to think of some *u'* ($\neq u$) which also expresses a correct

answer to *Q* and they will come to understand *q* by recognizing that *u'* expresses a correct answer to *Q*. In such a case I think we would say that *S* is not explaining *q* by uttering *u* but uttering something which he believes will cause others to be able to explain *q*. To preclude this possibility we might say that *S* intends his utterance of *u* to render *q* understandable by producing the recognition that *u* expresses a correct answer to *Q*, and *S* does not intend his utterance of *u* to render *q* understandable by producing the recognition that some *u'* expresses a correct answer to *Q*. Using the term "solely" to abbreviate the second conjunct of this condition we can write

- (3) *S* utters *u* with the intention that his utterance of *u* render *q* understandable solely by producing the recognition that *u* expresses a correct answer to *Q*.⁶

One further modification in this condition is needed whose expression requires concepts to be introduced in section 8. With this modification (see footnote 18) I am inclined to think that these three conditions are not only necessary but jointly sufficient for "*S* explains *q* by uttering *u*." If so then the same honor can be accorded to (3) by itself, since (3) entails both (1) and (2). However, even if they are necessary but not sufficient, I believe that they expose enough features of the act of explaining to enable us to consider the product of such acts.

Although "explain" may be used in describing an act governed by these conditions, it can also be employed in a more restricted way to cover only correct explainings. We can say that the doctor explained Bill's stomach ache, even though he did so incorrectly, or that no one has yet explained Bill's stomach ache, even though the doctor has tried. When one has correctly explained *q* by uttering *u* one has performed the illocutionary act of explaining *q* and in doing so one has provided a correct answer to *Q* (more about this in sections 10 and 12). In what follows, however, when reference is made to acts of explaining I shall mean acts for which this is not a requirement. The explanations whose ontology will be studied need not be correct ones.

⁴ For an analysis and defense of this condition see my *Law and Explanation* (Oxford, 1971), Chapter 4.

⁵ If '*q*' is an indirect interrogative, then *Q* is the question it expresses. Later I will suggest that any sentence of the form "*S* explains *q* by uttering *u*" is transformable into one in which '*q*' is an indirect interrogative. The present condition could be weakened to accommodate cases in which *S* believes that *u* expresses what *might* be a correct answer to *Q* (though he is unsure), and cases in which *S* believes that *u* expresses an answer that is approximately correct. See *Law and Explanation*, *ibid.*, pp. 63, 79.

⁶ I am indebted to participants in my explanation seminar (Fall, 1974), especially to Timothy McCarthy and George Wilson, for helping me see the need for a condition such as (3) with certain Gricean features.

3. *Product: Sentence and Proposition Views*

What kind of entity is the product of an act of explaining? If *S* explains *q* by uttering something, it might naturally be supposed that the product-expression in

- (1) The explanation of *q* given by *S* is that *p*
denotes what *S* utters, let us say a sentence. Assuming that the product-expression in (1) is uniquely referring we can formulate the following

Denotation condition (sentence view): "The explanation of *q* given by *S*" denotes *u* if and only if

- (i) *u* is a sentence
- (ii) *S* explained *q* by uttering *u*
- (iii) $(\forall v)[(S \text{ explained } q \text{ by uttering } v) \supset (v = u)]$.⁷

In accordance with this view explanations are sentences (including conjunctions) by the uttering of which explainers explain.

Two major problems beset the sentence theory, but since these are common to other views I shall deal with them later. However, a difficulty will now be mentioned which suggests the preferability of a proposition view. Suppose that Dr. Smith explained Bill's stomach ache by uttering

- (2) Bill ate spoiled meat,
while Dr. Robinson explained it by uttering
(3) Bill ate meat that was spoiled.

If (2) is the only sentence by the uttering of which Dr. Smith explained Bill's stomach ache, and (3) is the only sentence by the uttering of which Dr. Robinson did so, then, on the present view, "the explanation of Bill's stomach ache given by Dr. Smith" denotes sentence (2), while "the explanation of Bill's stomach ache given by Dr. Robinson" denotes sentence (3). But since sentence (2) $\neq$ sentence (3),

the explanation of Bill's stomach ache given by Dr. Smith $\neq$ the explanation of Bill's stomach ache given by Dr. Robinson,

which seems unsatisfactory. Intuitively, both doctors have given the same explanation. Both have

attributed the stomach ache to the eating of spoiled meat, despite the fact that the particular sentences used by each are not the same.

This multiplication of explanations can be avoided by identifying explanations with propositions, not sentences. Since sentences (2) and (3) express the same proposition we can conclude that the explanations given by the doctors are the same. Let us then formulate the following

Denotation condition (proposition view): "The explanation of *q* given by *S*" denotes *x* if and only if

- (i) *x* is a proposition
- (ii) $(\exists u)[S \text{ explained } q \text{ by uttering } u]$
- (iii) $(\forall u)[(S \text{ explained } q \text{ by uttering } u) \supset (u \text{ expresses } x)]$.

On this view explanations are propositions expressed by sentences by the uttering of which explainers explain.

Two considerations, however, make this (as well as the previous view) untenable.

4. *The Illocutionary Force Problem*

Suppose that by uttering the sentence

- (1) Bill ate spoiled meat
Jane criticized Bill for eating spoiled meat. Then
(2) The criticism of Bill given by Jane is that Bill ate spoiled meat.

By analogy with the proposition view of the product of explanation, the product-expression in (2) —"the criticism of Bill given by Jane"—will be taken to denote the proposition expressed by the sentence Jane uttered in giving her criticism, viz. the proposition expressed by (1).⁸ On the proposition view, then, the product-expression in (2) denotes the same proposition as the product-expression "the explanation of Bill's stomach ache given by Dr. Smith" in our earlier example. Therefore, the explanation of Bill's stomach ache given by Dr. Smith = the criticism of Bill given by Jane. But the criticism of Bill given by Jane was a criticism.

⁷ The product-expression in (1) may be used to uniquely refer even if *S* gave different explanations of *q*. E.g., when we use this expression we might be referring to the explanation of *q* given by *S* during time *t*, in which case the expression "S explained *q* by uttering *u*" in (ii) and (iii) becomes "S explained *q* by uttering *u* during *t*." Again the product-expression in (1) may be used to uniquely refer even if *S* explained *q* on many occasions provided he gave the same explanation on each.

Here and in what follows I will make the simplifying assumption that explainers always explain by uttering something rather than by performing some non-verbal act such as handwaving. To deal with explanations not produced by uttering something we could speak of "verbal equivalents" of such non-verbal acts and change conditions (ii) and (iii) by writing "S explained *q* by performing an act whose verbal equivalent is uttering *u*."

⁸ The denotation condition would be this: "the criticism given by *S*" denotes *x* iff (i) *x* is a proposition, (ii) $(\exists x)[S \text{ criticized by uttering } u]$, (iii) $(\forall u)[(S \text{ criticized by uttering } u) \supset (u \text{ expresses } x)]$.

Therefore,

- (3) The explanation of Bill's stomach ache given by Dr. Smith was a criticism.

Now one possible way of construing (3) is this:

Dr. Smith's act of explaining Bill's stomach ache was an act of criticizing,

which is false if Dr. Smith explained Bill's stomach ache but did not criticize anyone. But another way of construing (3)—the way needed for the present argument—is one according to which the expression "the explanation of Bill's stomach ache given by Dr. Smith" in (3) is supposed to denote the *product* of Dr. Smith's act of explaining, not that act itself. Is (3) also false under this interpretation?

I believe that it is. On the product view (3) is to be understood as having the form " $(\lambda x)Fx$ was a criticism," in which " $(\lambda x)Fx$ " denotes a product of explanation, and "was a criticism" is true of that product. Assuming that products of explanations are entities that can be given or made, from any sentence of the above form we ought to be able to infer "whoever gives $(\lambda x)Fx$ gives a criticism." Therefore, from (3) we may infer

Whoever gives the explanation of Bill's stomach ache given by Dr. Smith gives a criticism.

But Dr. Smith gave the explanation of Bill's stomach ache given by Dr. Smith. Therefore,

- (4) Dr. Smith gave a criticism,

which is false if Dr. Smith had no intention to be critical. What is true is only that what Jane uttered in criticizing Bill was what Dr. Smith uttered in explaining his stomach ache. But it does not follow from this that Dr. Smith gave a criticism.

This will be called the illocutionary force problem. The proposition expressed by what is uttered in an act of explaining may be the same as the proposition expressed by what is uttered in other illocutionary acts, such as an act of criticizing. If the product-expression denotes such a proposition then we will have to conclude that the explanation given by *S* is a criticism, and hence that when *S* gave an explanation he gave a criticism, even when this is not so. The proposition view (as well as the sentence view) is beset by the illocutionary force problem.

Since this charge is important in what follows I want to consider two replies a proposition theorist might be tempted to make. First, he might point out that "*S* gives the (or, a) criticism *y*" has been treated above as expressing a "giving" relationship between the critic and the (or some) product of his

act of criticizing. A proposition theorist may, however, want to understand "*S* gives the (a) criticism *y*" to mean "*S* performs an illocutionary act whose product is the (a) criticism *y*." Even so the illocutionary force problem arises. In the previous argument "whoever gives $(\lambda x)Fx$ gives a criticism" becomes: "whoever performs an illocutionary act whose product is $(\lambda x)Fx$ performs an illocutionary act whose product is a criticism." Therefore from (3) we may infer

Whoever performs an illocutionary act whose product is the explanation of Bill's stomach ache given by Dr. Smith performs an illocutionary act whose product is a criticism.

But Dr. Smith performed an illocutionary act whose product is the explanation of Bill's stomach ache given by Dr. Smith. Therefore,

Dr. Smith performed an illocutionary act whose product is a criticism.

But this, on the present proposal, means the same as (4), and (4) we are supposing to be false. We get the same result as before without assuming that "*S* gives a criticism *y*" expresses a "giving" relationship between *S* and some criticism.

Even so a proposition theorist may not be satisfied, and may propose an argument to show that, contrary to what we have been supposing, (3) and (4) are true, not false. (I will give the argument for truth of (3); an analogous one is possible for (4).) We start as we did by assuming (2) and

- (5) The explanation of Bill's stomach ache given by Dr. Smith is that Bill ate spoiled meat.

From these we obtain

- (6) The explanation of Bill's stomach ache given by Dr. Smith was a criticism of Bill given by Jane.

And from the latter we infer (3). So (3) is indeed true if we assume—as we have been doing—that (2) and (5) are true.

However, in this argument the step from (6) to (3) is unwarranted, as can be seen if we shift to the following example with the same form of inference. Mary gives answer *A* to question *Q*₁ and John also gives answer *A* but to a different question *Q*₂. Let John's answer to *Q*₂ be incorrect. We might then say

- (7) The answer given by Mary to *Q*₁ was an incorrect answer given by John to *Q*₂.

But from this we cannot conclude that the answer given by Mary to *Q*₁ was an incorrect answer. In

(7) the predicate "was an incorrect answer given by John to Q_2 " should not be construed as a conjunction of "was an incorrect answer" and "was given by John to Q_2 " but as "was incorrect as an answer given by John to Q_2 ." Similarly in (6) "was a criticism of Bill given by Jane" should be construed as "was a criticism of Bill as given by Jane." And from " x is F as a G " we cannot infer " x is F ."

5. The Emphasis Problem

The second problem with the proposition view is that its denotation condition assumes that the u -position in sentences of the form " S explained q by uttering u " is referentially transparent when this position is filled by expressions for sentences. But this assumption can be seen to be unjustified by appeal to the notion of emphasis.⁹

There are expressions for sentences which are such that adding or changing emphasis in them will not alter the sentence referred to. For example, suppose I refer to

- (1) The sentence "Bill ate spoiled meat on Tuesday,"

and you, who are hard of hearing, think I was referring to the sentence "Bill ate spoiled meat on Monday." I might reply that I am referring to

- (2) The sentence "Bill ate spoiled meat *on Tuesday*."

By using emphasis in (2) I am not referring to a different sentence but to the same one as in (1). Only I use emphasis in (2) and not in (1) to correct your mistake about which sentence I am referring to. In the paper cited in footnote 9 this is what I call a non-semantic use of emphasis. When such a use of emphasis occurs then neither the meaning nor the reference of the referring expression is altered with a change in emphasis. Accordingly, the sentence "Bill ate spoiled meat on Tuesday" = the sentence "Bill ate spoiled meat *on Tuesday*."¹⁰ Similarly, the sentence "Bill ate spoiled meat on Tuesday" = the sentence "Bill ate *spoiled meat* on Tuesday." Hence

- (3) The sentence "Bill ate *spoiled meat* on Tuesday" = the sentence "Bill ate spoiled meat *on Tuesday*."

Suppose now that

- (4) Dr. Smith explained Bill's stomach ache by uttering the sentence "Bill ate *spoiled meat* on Tuesday."

Dr. Smith emphasizes "spoiled meat" to indicate that he believes that this aspect of the situation is relevant for explaining Bill's stomach ache. He believes that it was the spoiled meat he ate that caused Bill's stomach ache. Now if the u -position in sentences of the form " S explained q by uttering u " is referentially transparent then from (4), in virtue of (3), we may infer

- (5) Dr. Smith explained Bill's stomach ache by uttering the sentence "Bill ate spoiled meat *on Tuesday*."

But (5) is false since it entails that Dr. Smith believes that the day on which Bill ate the spoiled meat was relevant for his getting a stomach ache. And I am supposing that Dr. Smith had no such belief. His claim was that Bill got a stomach ache because of what he ate and not because of when he ate it. This distinction is expressed by the difference in emphasis in (4) and (5). It is also expressed by the difference in emphasis between "the explanation of Bill's stomach ache given by Dr. Smith is that Bill ate *spoiled meat* on Tuesday," which is true if (4) obtains, and "the explanation of Bill's stomach ache given by Dr. Smith is that Bill ate spoiled meat *on Tuesday*," which is true if (5) obtains.

The emphases in (2) and

- (6) The sentence "Bill ate *spoiled meat* on Tuesday" are non-semantic. A shift of emphasis in these expressions will not change their meanings or references. But when these expressions are embedded in explanation-sentences of the form (4) the emphases can assume a semantic role. The emphasized words can become "captured" by the term "explained," indicating that a particular aspect of Bill's eating spoiled meat on Tuesday is claimed to be explanatorily relevant for his getting a stomach ache. A shift in emphasis in (6) as this appears in (4) can transform (4), which is true, into (5), which is false.

This is not to deny that the emphases in both (4) and (5) could be used to play non-semantic roles. (One might assert (5) in response to someone who

⁹ For a general discussion of emphasis and reference that can be used to support the claims about to be made, see my "Causation, Transparency, and Emphasis," *Canadian Journal of Philosophy*, vol. 5 (1975), pp. 1-23. For a seminal discussion of emphases see Fred Dretske, "Contrastive Statements," *The Philosophical Review*, vol. 81 (1972), pp. 411-437.

¹⁰ We are, of course, speaking here of sentence types, not tokens.

thinks that Dr. Smith explained Bill's stomach ache by uttering the sentence "Bill ate spoiled meat on Monday.") But if so (4) and (5) would have different readings from the ones being given here. When emphasis is understood as playing a semantical role—which is possible in (4) and (5)—we get readings for these sentences under which (4) is true and (5) false.

Assuming, then, that (2) and (6) denote the same sentence, and that the substitution of (2) for (6) in (4) turns a true sentence into a false one, we must conclude that the *u*-position in sentences of the form "*S* explained *q* by uttering *u*" is referentially opaque if that position is filled by sentence-expressions. There are readings for sentences of this form such that the substitution of co-referring *u*-terms will lead from truths to falsehoods. But the denotation condition of the proposition (as well as the sentence) view assumes that the *u*-position is transparent.

This problem can be avoided by introducing the notion of an *e*-sentence ('*e*' for emphasis), which is a sentence together with its emphasis, if any. To give an *e*-sentence that *S* uttered in explaining *q*, one must supply *S*'s words, in the order he gave them, and with any emphasis he used. And we will now suppose that what is uttered by speakers in acts of explaining are *e*-sentences. The problem above is thus avoided since

- (7) The *e*-sentence "Bill ate *spoiled meat* on Tuesday"

is not identical with

- (8) The *e*-sentence "Bill ate spoiled meat on Tuesday."

Therefore from the fact that Dr. Smith explained Bill's stomach ache by uttering the *e*-sentence denoted by (7) it does not follow that he explained Bill's stomach ache by uttering the *e*-sentence denoted by (8).

Nevertheless there is a residual emphasis problem for the proposition view, which must now assume that *e*-sentences express propositions. Although the *e*-sentences denoted by (7) and (8) are not identical, the propositions these *e*-sentences express are. (7) and (8), respectively, denote the *e*-sentences

- (9) Bill ate *spoiled meat* on Tuesday,

- (10) Bill ate spoiled meat on Tuesday.

A shift in emphasis in (9) or (10) changes the *e*-sentence but not the proposition expressed. If I utter the *e*-sentence (9)—using emphasis where I do to in-

dicate my surprise—and you think I am expressing the proposition that Bill ate spoiled meat on Monday, I may correct your mistake by uttering (10). By doing so I am not expressing a different proposition but the original one in a manner that will correct your misconception (i.e., by using a different *e*-sentence).

Suppose then that Dr. Smith explained Bill's stomach ache by uttering the *e*-sentence (9), while Dr. Jones explained it by uttering the *e*-sentence (10). And in both cases assume that requirement (iii)—the uniqueness requirement—of the proposition theory's denotation condition is satisfied. It follows from this denotation condition that "the explanation of Bill's stomach ache given by Dr. Smith" denotes the same proposition as "the explanation of Bill's stomach ache given by Dr. Jones," viz. the proposition expressed by the *e*-sentences (9) and (10). Therefore

- (11) The explanation of Bill's stomach ache given by Dr. Smith = the explanation of Bill's stomach ache given by Dr. Jones,

which we ought to reject since according to Dr. Jones' explanation, but not Dr. Smith's, the date on which Bill ate the spoiled meat is relevant for explaining Bill's stomach ache. If the date was irrelevant—if Bill would have gotten a stomach ache from eating spoiled meat on any day of the week—then, we may suppose, Dr. Smith's explanation is correct while Dr. Jones' is incorrect. We must conclude that these explanations are different and hence that (11) is false.

6. The Argument View

Before pursuing a remedy for these problems let me turn briefly to a different product view, suggested by the D-N model, according to which explanations are arguments.

Denotation condition (Argument view): "The explanation of *q* given by *S*" denotes an argument one of whose premisses is (the proposition expressed by) *u* if and only if

- (i) *S* explained *q* by uttering *u*
 (ii) $(\forall v)[(S \text{ explained } q \text{ by uttering } v) \supset (v = u) \text{ (or } v \text{ and } u \text{ express the same proposition)}]$.

On one version of this view arguments are composed of propositions, and on the other of sentences. Since the proposition version will allow us to count explanations as identical even if they use different sentences, it will be considered. To insure referential transparency in the *u*-position in "*S* explained *q* by uttering *u*" let us assume that the *u*'s are *e*-sentences.

Suppose, then, that Dr. Smith explained Bill's stomach ache by uttering the *e*-sentence

- (1) Bill ate *spoiled meat* on Tuesday.

Assuming that condition (ii) above is satisfied the product-expression in

- (2) The explanation of Bill's stomach ache given by Dr. Smith is that Bill ate *spoiled meat* on Tuesday

denotes some argument one of whose premisses is the proposition expressed by (1). What argument?

Following the D-N model we may assume that it contains a law among its premisses and that its conclusion, which is entailed by the premisses, describes the event referred to in the object-expression. We might suppose that the product-expression in (2) denotes an argument composed of propositions expressed by sentences such as these:

- (3) Bill ate *spoiled meat* on Tuesday. Anyone who eats spoiled meat gets a stomach ache. Hence, Bill got a stomach ache.

On the D-N view an argument such as (3) is to be construed as an ordered set of propositions (or sentences), or perhaps better, as an ordered pair whose first member is the conjunction of premisses and whose second is the conclusion.

The argument view fares no better than the proposition view of section 3. First, there is an illocutionary force problem. Poor Sam dies one hour after being operated on by Dr. Smith, who offers an excuse by uttering

- (4) Sam had disease *d* at the time of his operation. Anyone with disease *d* at the time of an operation dies one hour after the operation. Hence, Sam died one hour after his operation.

Suppose further that Dr. Jones explains Sam's death by uttering (4). I will assume that on the argument view of product, both "the excuse given by Dr. Smith" and "the explanation of Sam's death given by Dr. Jones" denote the argument given by (4), i.e., the ordered pair whose first proposition is expressed by the conjunction of the first two sentences in (4) and whose second proposition is expressed by the last sentence. If so then the excuse given by Dr. Smith = the explanation of Sam's death given by Dr. Jones. But the excuse given by Dr. Smith was an excuse. Therefore,

- (5) The explanation given by Dr. Jones was an excuse,

which is false if Dr. Jones had no intention of excusing anything. Indeed Dr. Jones may believe the operation and resulting death to be inexcusable. What is true is only that what Dr. Smith uttered in giving his excuse was what Dr. Jones uttered in explaining Sam's death. But (5) does not follow from that.

The emphasis problem can also be shown to be present, but I will omit the details.

7. Propositions Construed as Explanatory

A view will now be developed which avoids the illocutionary force problem. Later it will be shown how it can be modified to avoid the emphasis problem as well.

The first difficulty with the proposition (and the argument) view is that *S* may explain *T* by uttering what is uttered by someone performing a different type of illocutionary act, such as criticizing or excusing. If product-expressions denote propositions then *S*'s explanation will turn out to be a criticism or an excuse even when it is not. The proposition view fails to take proper account of the fact that to explain is to utter something with a certain illocutionary force. Perhaps then the product of an explanation is what is expressed by that utterance—say a proposition—construed with the illocutionary force of explaining.

Searle in a discussion of illocutionary acts criticizes Austin for failing to distinguish two senses of "statement": the act of stating and what is stated (which Searle calls the statement-object and I call the product of the act of stating).¹¹ He goes on to identify the latter as a "proposition construed as stated."¹² His discussion here is very brief and we are not told what sort of thing this is. But if we knew we could better understand product of explanation. That is, following Searle's discussion of statements, having distinguished the act of explaining from the product of the act, we now identify the latter as a proposition construed as explaining. But what sort of entity is that?

The proposal I shall now consider is that it is an ordered pair consisting of a proposition and a type of illocutionary act, in the present case an explaining type of act. So, e.g.,

- (1) (the proposition that Bill ate *spoiled meat*, explaining)

is a product of explanation, since it is an ordered pair whose first member is a proposition and whose

¹¹ John Searle, "Austin on Locutionary and Illocutionary Acts," *The Philosophical Review*, vol. 77 (1968), pp. 305-424.

¹² *Ibid.*, p. 423.

second is the act-type explaining. In the sentence

- (2) The explanation of Bill's stomach ache given by Dr. Smith is that Bill ate spoiled meat

the product-expression could be said to denote the ordered pair (1) provided that Dr. Smith explained Bill's stomach ache by uttering something and whatever the doctor explained Bill's stomach ache by uttering expresses the proposition that Bill ate spoiled meat. More generally, we have the following

Denotation condition (Ordered pair view): "The explanation of q given by S " denotes (x, y) if and only if

- (i) x is a proposition
- (ii) $y =$ (the act-type) explaining
- (iii) $(\exists u)(S \text{ explained } q \text{ by uttering } u)$
- (iv) $(\forall u)[(S \text{ explained } q \text{ by uttering } u) \supset (u \text{ expresses } x)]$.¹³

It is important now to distinguish two uses of explanation product-expressions. We often use "the explanation of q given by S " so that a change in the object of explanation (the q) need not mean a change in the product. In the language of the D-N model, the term "explanation" can be used to refer to the explanans without the explanandum. This will be called the "partial sense" of explanation. It is the sense intended by the denotation condition above. According to this condition, "the explanation of Bill's stomach ache given by Dr. Smith" and "the explanation of Bill's fever given by Dr. Smith" could denote the same entity, viz. the ordered pair (1), provided that the doctor explained both Bill's stomach ache and his fever, and whatever he explained these by uttering expresses the proposition that Bill ate spoiled meat.

However, an expression of the form "the explanation of q given by S " is also used in such a way that a change in the object of explanation means a change in the product. In the language of the D-N model, an explanation consists of both explanans and explanandum. Let me refer to this as the "full sense" of explanation. In accordance with this usage, the explanation of Bill's stomach ache given by Dr. Smith could not be the same entity as the explanation of Bill's fever given by Dr. Smith, since different things are being explained. The ordered pair view can accommodate this "full sense" by including the object of explanation q in the type of act. In the denotation condition above (ii) is changed to " $y =$ explaining q ."

The ordered pair view avoids the illocutionary force problem, as can be seen if we consider just the "full sense" version. The product-expression in (2) denotes the ordered pair.

- (3) (the proposition that Bill ate spoiled meat, explaining Bill's stomach ache),

while that

- (4) The criticism of Bill given by Jane is that Bill ate spoiled meat

denotes

- (5) (the proposition that Bill ate spoiled meat, criticizing Bill).

But (3) $\neq$ (5). Therefore, neither "the explanation of Bill's stomach ache given by Dr. Smith = the criticism of Bill given by Jane" nor "the explanation of Bill's stomach ache given by Dr. Smith was a criticism" can be derived. This view avoids the illocutionary force problem by taking into account not simply the proposition expressed in an act but also the illocutionary act itself. In (2) and (4) the types of illocutionary acts are different even though the propositions expressed are the same.

On the other hand, the emphasis problem remains, as can be seen if we suppose that Dr. Smith explained Bill's stomach ache by uttering "Bill ate spoiled meat on Tuesday," while Dr. Jones explained it by uttering "Bill ate spoiled meat on Tuesday." On the "full sense" version the explanation given by each doctor is the ordered pair (the proposition that Bill ate spoiled meat on Tuesday, explaining Bill's stomach ache), since the e -sentences uttered by doctors express the same proposition. Hence, the explanation of Bill's stomach ache given by Dr. Smith = the explanation of Bill's stomach ache given by Dr. Jones, which is false.

Can the present view be modified to avoid this problem? In the following section a proposal will be suggested for the full-sense version, although an analogous one is possible for the partial-sense version.

8. A New Ordered Pair View

In section 2 among the conditions for S performing an explaining act is that S believes that what he utters expresses a correct answer to some question Q . A view of explanations as answers to questions will now be explored. I will assume that such questions are *wh*-questions (those expressed using interrogative pronouns such as "why," "what,"

¹³ Analogous conditions are possible for other illocutionary product-expressions such as "the criticism of q given by S ."

"how," "who," "when," and "where"—but not "whether" questions, which call for a "yes" or "no" answer).

A question such as

(1) Why did Bill get a stomach ache?

will be said to *presuppose* a number of propositions, e.g.,

(2) Bill got a stomach ache for some reason

Bill got a stomach ache

Bill got something.¹⁴

Any proposition entailed by a proposition presupposed by a question will also be said to be presupposed by that question. A *complete presupposition* of a question is a proposition that entails all and only the presuppositions of that question. Of the propositions above only (2) is a complete presupposition of (1).

To obtain a complete presupposition of a (*wh*-) question it will be necessary to use some *existential* term or phrase which corresponds to the interrogative pronoun in a sentence expressing that question. Thus "for some reason" in (2) corresponds to "why" in (1). (In the case of other questions "in some manner" could correspond to "how," "at some time" to "when," and so forth.) A sentence expressing a proposition that is a complete presupposition of question *Q* can be transformed into one expressing a proposition that is a *complete answer* to *Q* by replacing that existential term by a phrase giving a reason, manner, time, etc. For example, the propositions expressed by the sentences

(3) Bill got a stomach ache because he ate spoiled meat

(4) Bill got a stomach ache because of something he ate

are complete answers to question (1).

"For some reason" will be called a reason-existential term ("at some time" a time-existential term, etc.). More generally, I shall speak of a ϕ -existential term and say that such a term can be replaced by a phrase giving a ϕ . Although I shall

not try to propose general conditions for saying when a phrase gives a reason, time, etc., I will say that a proposition *p* is a complete answer to *wh*-question *Q* if and only if

- (i) there is a sentence S_q that expresses *Q* and begins with an interrogative pronoun *i*;
- (ii) there is a sentence *S* which expresses a proposition that is a complete presupposition of *Q*, and *S* contains a ϕ -existential term which corresponds to the interrogative pronoun *i*;
- (iii) *S* can be transformed into a sentence S_p which expresses proposition *p* by replacing the ϕ -existential term by a phrase giving a ϕ ;
- (vi) there is no mutual entailment between *S* and S_p .¹⁵

By this definition propositions (3) and (4) are complete answers to (1) because sentences expressing these propositions can be obtained from sentence (2) by replacing the reason-existential term in (2) by some phrase giving a reason. Any complete answer to *Q* will entail all the presuppositions of *Q*.¹⁶ Accordingly, the proposition that Bill ate spoiled meat is not a complete answer to (1)—though it is an answer—because it does not entail the proposition that Bill got a stomach ache, which is a presupposition of (1).

Returning now to the ordered pair view, my proposal is that the constituent proposition in the ordered pair is a complete answer to a certain question. This question will be given by the object-expression. It will be supposed that any product-expression of the form "the explanation of *q* given by *S*" is or is transformable into one in which '*q*' is an indirect interrogative whose direct form expresses a question *Q*. Thus the product-expression "the explanation of Bill's stomach ache given by Dr. Smith" could be reformulated as "the explanation given by Dr. Smith of why Bill got a stomach ache." And the latter will be taken to denote an ordered pair one of whose constituents is a proposition that is a complete answer to question (1). But which proposition will this be and how is it related to Dr. Smith's act of explaining?

¹⁴ Questions are here being treated as abstract entities on a par with propositions. Nuel Belnap proposes a definition of presupposition for questions, as follows: question *Q* presupposes proposition *p* if and only if the truth of *p* is a logically necessary condition for there being some true answer to *Q*. "Questions: Their Presuppositions and How they Can Fail to Arise" in Karel Lambert (ed.), *The Logical Way of Doing Things* (New Haven, 1969), pp. 23–37.

¹⁵ Condition (iv) precludes "Bill got a stomach ache for the reason that Bill got a stomach ache" from expressing a complete answer to (1). An even stronger condition would be to require that the phrase giving a ϕ , if a singular term (or if transformed into one), be a rigid designator. This would preclude not only the above example but examples such as "Bill got a stomach ache for the reason that Sam did."

¹⁶ What I am calling a complete answer to *Q* may, of course, not be very complete at all in the usual sense of this term. But it is complete in the sense of entailing all the presuppositions of *Q*.

Suppose there is an act in which *S* explained *q* by uttering *u*. (Henceforth '*q*' will always be an indirect interrogative expressing a question *Q*.) If *p* is a proposition satisfying the following conditions I shall say that *p* is *associated with* that act of explaining:

- (a) *p* is a complete answer to *Q*
- (b) the act was one in which *p* was claimed to be true
- (c) *p* entails the proposition expressed by *u*.

According to these conditions, if Dr. Smith explained why Bill got a stomach ache by uttering

- (5) Bill got a stomach ache because he ate *spoiled meat* on Tuesday,

then the proposition (5) is associated with the act of explaining performed by Dr. Smith. But equally if Dr. Smith explained why Bill got a stomach ache by uttering

- (6) Bill ate *spoiled meat* on Tuesday,

then proposition (5) is also associated with Dr. Smith's act of explaining, even though proposition (5) is not expressed by the *e*-sentence which Dr. Smith uttered. If by uttering (6) Dr. Smith explained why Bill got a stomach ache, his act of explaining was one in which the truth of proposition (5) was still being claimed. Since proposition (5) is a complete answer to question (1) and entails the proposition expressed by sentence (6) which Dr. Smith uttered, proposition (5) is associated with his act of explaining.

Now the constituent proposition in the explanation of *q* given by *S* will be one *associated with S's* act(s) of explaining *q*. Accordingly, we may formulate the following

Denotation condition (New ordered pair view): "the explanation of *q* given by *S*" denotes (*x*, *y*) if and only if

- (i) *x* is a proposition that is a complete answer to *Q*
- (ii) *y* = explaining *q*
- (iii) $(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u)$
- (iv) $(\forall a)(\forall u)[(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u) \supset (\forall r)(r \text{ is associated with } a \equiv r = x)]$.¹⁷

Since proposition (5) is a complete answer to question (1), "the explanation given by Dr. Smith of why Bill got a stomach ache" denotes (proposi-

tion (5), explaining why Bill got a stomach ache), provided that there was an act in which Dr. Smith by uttering something explained why Bill got a stomach ache, and for any such act proposition (5) was the one and only associated proposition.

Returning now to the definition of association, why are such propositions and only these taken to be constituent propositions of explanations? That is, why are conditions (a)–(c) imposed? The first, by requiring that *p* be an answer to *Q*, relates the product of explanation to the act in which the explainer intends to provide an answer to *Q*; and by requiring that answer to be complete the emphasis problem is avoided (as shown in the next section). The reason for condition (b) can be shown by an example in which Dr. Smith by uttering the *e*-sentence (6) explains why Bill got a stomach ache. The proposition

- (7) Bill got a stomach ache because Bill ate spoiled meat *on Tuesday*

is a complete answer to question (1), and it entails the proposition expressed by the *e*-sentence Dr. Smith uttered, viz. (6). But (7) could not be the constituent proposition in Dr. Smith's explanation (otherwise his explanation would be identical with Dr. Jones', which it isn't). Condition (b) precludes this on the ground that (7) is not a proposition claimed to be true in Dr. Smith's act of explaining. By uttering (6) in explaining why Bill got a stomach ache Dr. Smith does not claim the truth of (7).

The reason for condition (c) can also be given by means of the above example in which Dr. Smith by uttering (6) explained why Bill got a stomach ache. The proposition

- (8) Bill got a stomach ache because he ate spoiled meat

is a complete answer to question (1), and it is also a proposition claimed to be true in Dr. Smith's act of explaining. Both propositions (5) and (8) satisfy conditions (a) and (b), and without condition (c) there would be no unique proposition to associate with Dr. Smith's explaining act. Condition (c) precludes (8) as a proposition associated with this act since (8) does not entail the proposition expressed by the *e*-sentence Dr. Smith uttered, viz. (6).

In choosing the constituent proposition to be one satisfying the conditions of *association* the product of explanation is seen to be intimately related to the act of explaining. The constituent proposition is an

¹⁷ Given the definition of "association" (i) is redundant since it is entailed by (iii) and (iv).

answer to a question (a question which the explainer in his act intends to answer); it is one claimed to be true in the explaining act; and either it is one expressed by what the explainer utters in that act or it entails the proposition expressed by what the explainer utters.¹⁸

The new ordered pair view can also supply conditions for being an explanation of q given by S . Here since uniqueness is not required, we need not suppose that the same proposition is associated with every act in which S explained q . Accordingly, we can write

(x, y) is an explanation of q given by S if and only if

- (i) x is a proposition that is a complete answer to Q
- (ii) $y = \text{explaining } q$
- (iii) $(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u \text{ and } x \text{ is associated with } a)$.

The ordered pair view can be extended, as shown in the appendices. What now needs to be asked, however, is whether it is subject to the difficulties of the previous views.

9. Are The Illocutionary Force and Emphasis Problems Avoided?

Since the second constituent of the product of explanation is an explaining type of act we avoid the illocutionary force problem. Even if the constituent proposition of Dr. Smith's explanation is the same as that of Jane's criticism it will not follow that "Dr. Smith's explanation of why Bill got a stomach ache" and "Jane's criticism of Bill" denote the same entity.¹⁹

Secondly, if the constituent proposition of an explanation is a complete answer to Q then the emphasis objection is avoided. Suppose that the explanation given by Dr. Smith of why Bill got a stomach ache is that Bill ate *spoiled meat* on Tuesday, and that the explanation given by Dr. Jones is that Bill ate spoiled meat *on Tuesday*. The emphasis

problem arises if we take the constituent propositions of these explanations to be ones expressed by the e -sentences

- (1) Bill ate *spoiled meat* on Tuesday
- (2) Bill ate spoiled meat *on Tuesday*.

These propositions are identical, from which we would have to conclude that the explanations given by the two doctors are identical. However, the propositions expressed by (1) and (2) are not complete answers to "why did Bill get a stomach ache?" Therefore, on the ordered pair view of section 8 these propositions are not the constituent propositions of the explanations in question. But we can suppose that the constituent proposition of the explanation given by Dr. Smith is

- (3) Bill got a stomach ache because he ate *spoiled meat* on Tuesday,

while that of the explanation given by Dr. Jones is

- (4) Bill got a stomach ache because he ate spoiled meat *on Tuesday*.

And these propositions are not identical. Shifting emphasis in (3) and (4), unlike (1) and (2), means changing the propositions expressed. If Dr. Smith by uttering (1) explained why Bill got a stomach ache, he put the emphasis where he did to indicate an explanatorily relevant aspect of an event. But the emphasis in (1) does not indicate this. By contrast, the emphasis in (3) does, since it is captured by "because." In general, in a complete answer any explanatory emphasis used by the explainer will be so captured.

10. Correct Explanations

None of the ontological views that have been discussed—neither the proposition, nor the argument, nor the ordered pair view—is committed by definition to some particular criterion for evaluating explanations. However, proponents of each view may have their favorite criterion and a given

¹⁸ We can now provide a more adequate formulation of condition (3), section 2, for " S explains q by uttering x ," as follows: S utters u with the intention that his utterance render q understandable either by producing the recognition that u expresses a correct answer to Q or by producing the recognition with respect to some u^* that it expresses a correct answer to Q , where u^* expresses a proposition that is a complete answer to Q , that S claims to be true when he utters u , and that entails the proposition expressed by u ; and S does not utter u with the intention that his utterance render q understandable by producing the recognition that some u' ($\neq u$ or u^*) expresses a correct answer to Q . This formulation, unlike the earlier one, permits S to intend his utterance of u to render q understandable in either of two ways; the second of these is by producing the recognition with respect to something other than u that it expresses a correct answer to Q , provided that it expresses a proposition that is associated with his explaining act.

¹⁹ It might be thought that if the constituent proposition in an explanation of q is a complete answer to Q then we can simply declare this proposition to be the product of explanation and dispense with the explaining type of act. But this will not avoid the illocutionary force problem. Jane's criticism of Bill might be that he got a stomach ache from eating spoiled meat. On the view that the product of explanation is a proposition that is a complete answer to Q , "Jane's criticism of Bill" and "Dr. Smith's explanation of why Bill got a stomach ache" would denote the same entity.

one might be said to be "natural" for or suggested by an ontological view.

A proposition theorist might say that since an explanation is a proposition, the question of whether an explanation is correct reduces to whether the proposition with which it is identical is correct or true:

- (1) (Possible proposition view): If proposition p is the explanation of q given by S then the latter is a correct explanation iff p is true.

Analogously, the argument and the ordered pair theorists might offer the following criteria:

- (2) (Possible argument view): If argument A is the explanation of q given by S then the latter is a correct explanation iff A is a correct argument (i.e., its premisses are true and entail its conclusion).²⁰
- (3) (Possible ordered pair view): If $(p, \text{explaining } q)$ is the explanation of q given by S , then the latter is a correct explanation iff p is a correct answer to Q .

Although (1)–(3) may seem natural, given respective ontologies, (1) and (2) are readily shown to be unreasonable in a way that (3) is not. Suppose that the coroner by uttering

- (4) The victim drank a deadly poison

explained how the victim died. Assume also that proposition (4) is true, but that drinking the deadly poison is not what caused the victim's death. Rather it was caused by a sudden, unrelated heart attack. Accordingly, the explanation given by the coroner is not a correct explanation. On the proposition view, the explanation given by the coroner = proposition (4). Since proposition (4) is, by hypothesis, true, we would have to conclude from (1) that the coroner's explanation is a correct explanation, which it isn't. On the argument view, the explanation given by the coroner is to be identified as an argument such as

The victim drank a deadly poison. Anyone who drinks a deadly poison dies. Hence, the victim died, in which proposition (4), which the coroner expressed, is a premiss. But this argument is correct (its premisses are true and entail its conclusion), and therefore by (2) the coroner's explanation is a correct explanation, which it isn't. Criteria (1) and (2) in conjunction with their respective ontological

theories of explanation lead to a difficulty which can be avoided only by proposing different criteria.

Criterion (3) does not have the same fate. For one thing, proposition (4) is not a complete answer to

- (5) How did the victim die?

Therefore the ordered pair view (proposition (4), explaining how the victim died) is not the explanation given by the coroner of how the victim died. Consequently criterion (3) cannot be applied to this ordered pair. For another thing, proposition (4) although true is not a correct answer to (5).

Even though an ordered pair theorist is not committed to adopting (3) in virtue of his ontology (in section 12 a different criterion will be noted), this may well be a reasonable criterion for him to adopt; but I will not pursue the matter here, except to make one observation. If it is adopted we cannot conclude that an explanation satisfying it is necessarily acceptable. An explanation may be correct, as far as it goes, but useless, since it does not go far enough. (In accordance with (3) someone's explanation of how the victim died may be correct even if its constituent proposition is that the victim died of natural causes.) Correctness is only one factor in evaluating an explanation; one must also consider the amount and type of the information sought. We want our explanations to be correct, but in an appropriate way. What counts as appropriate depends on contextual features concerning the knowledge and interests of those to whom the explanation is given.²¹

11. A No-Product View

A basic assumption of our discussion has been that explanation product-expressions denote entities. The only view of such entities that escapes previous objections is the ordered pair view of section 8. But an ontological purist may wonder whether he needs to postulate ordered pairs, or indeed anything else, as referents of explanation product-expressions. He may ask whether a sentence containing "the explanation of q given by S " can be understood without presupposing such entities but only particular acts of explaining. In this section a "reductionist" view will be considered according to which

Any sentence containing an explanation product-

²⁰ Where the premisses of A contain at least one law and the conclusion is a description of the event to be explained, (2) becomes a D-N view.

²¹ For a discussion of such features see my *Law and Explanation*, *op. cit.*, Chapter 4.

expression is paraphrasable into a sentence that contains no such expression, but does contain one or more expressions of the form "a is an act in which S explained q by uttering u."

When this thesis is combined with the view that such paraphrases are more fundamental than sentences with product-expressions we get the no-product view.

To see how the paraphrases are supposed to work let me consider sentences of the form

(1) The explanation of q given by S is that p.

I will say that such a sentence is *restructured* if 'q' is an indirect interrogative expressing a question Q and 'p' is a sentence expressing a complete answer to Q. It will be assumed that any unrestructured sentence of form (1) is paraphrasable into a restructured sentence of this form. A restructured sentence of form (1) can then be understood as

(2) $(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u) \text{ and } (\forall a)(\forall u)[a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u \supset (\forall r)(r \text{ is associated with } a \equiv r = p)]$,

i.e., there is an act in which S explained q by uttering something and for any such act p is the one and only associated proposition. To understand (2) we need to suppose that S performed one or more explaining acts, but no products of those acts need to be postulated.

A restructured sentence of the form "an explanation of q given by S is that p" can be paraphrased simply as " $(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u \text{ and } p \text{ is associated with } a)$," in which no product of explanation is invoked.

On the ordered pair view a sentence of the form

(3) The explanation of q given by S_1 = the explanation of q given by S_2

is to be understood as a genuine identity in which the expression on the left is being said to denote the same product as the one on the right. On the no-product view (3) is not a genuine identity but is to be understood as

$(\exists a)(\exists b)(\exists u_1)(\exists u_2)(a \text{ is an act in which } S_1 \text{ explained } q \text{ by uttering } u_1 \text{ and } b \text{ is an act in which } S_2 \text{ explained } q \text{ by uttering } u_2) \text{ and } (\exists p)(\forall a)(\forall b)(\forall u_1)(\forall u_2)[(a \text{ is an act in which } S_1 \text{ explained } q \text{ by uttering } u_1 \supset p \text{ is the one and only proposition associated with } a) \text{ and } (b \text{ is an act in which } S_2 \text{ explained } q \text{ by uttering } u_2 \supset p \text{ is the one and only proposition associated with } b)]$.

Finally, a sentence such as "the explanation of q given by S is correct" can be understood as

$(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u) \text{ and } (\exists p)(\forall a)(\forall u)[(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u) \supset (p \text{ is the one and only proposition associated with } a \text{ and } p \models \text{a correct answer to } Q)]$.

That is, there is an act in which S explained q by uttering something, and there is a proposition uniquely associated with all such acts, and that proposition is a correct answer to Q.

The present no-product view and the ordered pair view of section 8 differ over whether explanations are entities. According to the former a restructured sentence of the form (1) is to be understood as a conjunction of an existentially general and a universally general sentence describing acts of explaining, viz. (2). According to the latter, (1) is a singular sentence in which the product-expression denotes an ordered pair. Nevertheless, there is an interesting similarity between these views. On the ordered pair view a restructured sentence of form (1) will be true if and only if the product-expression in (1) denotes an ordered pair consisting of p and explaining q and (2) above obtains. On the no-product view a restructured sentence of form (1) is paraphrased into (2). So what is a paraphrase of (1) on the no-product view $\models$ one of the truth-conditions for (1) on the ordered pair view. And this is so not only for (1) but for other sentences that contain product-expressions.

Is one of these views superior? One consideration is how well each can justify inferences that are drawn from sentences containing explanation product-expressions. For example, from a sentence of form (1) we may quite naturally infer "S explained q by uttering something." Such an inference is sanctioned by the no-product view which paraphrases (1) as (2), which entails

(4) $(\exists a)(\exists u)(a \text{ is an act in which } S \text{ explained } q \text{ by uttering } u)$.

But the ordered pair view is equally successful, since (4) is derivable from (1) via the denotation condition for "the explanation of q given by S." More generally, it can be shown that for every inference sanctioned by one view there is an identical or closely corresponding one sanctioned by the other.

Is one of these views preferable on ontological grounds? Both the ordered pair and no-product views invoke acts of explaining and propositions. But the former in addition requires types of acts and ordered pairs consisting of these and propositions. Ontologists who crave simplicity may give

an edge to the no-product view, provided that paraphrases such as those above are adequate. On the other hand, a more liberal ontological position is possible. Two views may be equally satisfactory, even if one invokes more entities than the other, provided there are no special problems with these entities. Liberal ontologists who find types of acts no more problematic than propositions, and who have no reasons for rejecting ordered pairs, may well conclude that there is little to choose between the ordered pair and no-product views.

12. Implications for Standard Theories of Explanation

If one holds either of these views must a theory of explanation such as the D-N model be rejected? According to the latter the product-expression in "the explanation of q given by S is that p " denotes an entity, viz. a deductive argument. As previously indicated, this is unacceptable. But can the spirit of the theory be retained within a more successful product view?

It can, provided that the D-N theory is now understood as a theory not about the ontological character of explanations but only about their *correctness*. On this construction the D-N theorist could accept the denotation condition of the ordered pair view of section 8, but he would replace correctness condition (3) of section 10 with one requiring laws and deductive entailment. And he would restrict such a condition to certain explanations only, e.g., to those whose constituent propositions are expressible by sentences of the form " x because y ." (He would not be providing correctness conditions for explanations of how, where, or when . . .) If so we can formulate the following

- (1) *D-N correctness condition* (Ordered pair view): If (p , explaining q) is an explanation in which proposition p is expressible by a sentence consisting of "because" flanked by x and y , where x and y are both sentences, then (p , explaining q) is a correct explanation if and only if (i) y is a true sentence; (ii) y is a conjunction at least one of whose conjuncts is a law; (iii) y entails x .

However, once we interpret the D-N theory as one with no particular ontological commitments but only as a theory of correctness, then it is also possible to infuse a D-N spirit into the no-product view:

- (2) *D-N correctness condition* (No-product view): The explanation of q given by S is correct if

and only if $(\exists a)(\exists u)(a$ is an act in which S explained q by uttering u) and $(\exists p)(\forall a)(\forall u)(\exists x)(\exists y)[a$ is an act in which S explained q by uttering $u \supset p$ is the one and only proposition associated with a , and x and y are sentences, and the sentence consisting of "because" flanked by x and y expresses p , and y is a conjunction at least one of whose conjuncts is a law, and y entails x].

Thus the D-N theory, which as ordinarily expressed does have ontological commitments, can be reformulated so as to be neutral between product and no-product views. (This is not to say, of course, that (1) and (2) are acceptable criteria of correctness; but that is another matter.) More generally, there are theories of explanation (Salmon's statistical theory is another example)²² which express an ontological position about explanation. (Salmon views explanations as sets of sentences that include probability laws.) Although this ontological position may be objectionable these theories can be reformulated, without commitment to their original ontologies, as providing only correctness conditions. And the latter can be expressed within the framework of a product as well as of a no-product view. On the ordered pair view, e.g., the generalized form of such a correctness condition would be: If (p , explaining q) is an explanation, it is a correct explanation if and only if proposition p satisfies condition C. Theories providing various criteria of correctness would be distinguished by the conditions they impose on the proposition p .

APPENDIX A

PROOF-EXPLANATIONS

The ordered pair view of section 8 can be further developed in ways briefly discussed in this appendix and the next. It is also possible to extend the no-product view of section 11 by analogous procedures.

A scientist might explain why gases obey the ideal gas law by uttering simply

- (1) Gases obey the ideal gas law because they are composed of molecules satisfying Newton's laws of motion.

On the ordered pair view his explanation would be (proposition (1), explaining why gases obey the ideal gas law). But the scientist might also explain

²² Wesley C. Salmon, *Statistical Explanation and Statistical Relevance* (Pittsburgh, 1971).

this by making utterances in addition to (1) and in so doing produce a proof to show that anything composed of gas molecules that satisfy Newton's laws obeys the ideal gas law. When *S* explains *q* by uttering *u* and in so doing produces a proof, I shall say that *S*'s explanation is a proof-explanation. The argument view of section 6 in effect makes all explanations proof-explanations. But not all are, even in science. A scientist may explain *q* simply by asserting the truth of one or more propositions without producing any proof or demonstration. A proof-explanation is a product of an act of explaining in which a proof or demonstration is produced by the explainer as a feature of his illocutionary act. When Dr. Smith explained Bill's stomach ache by asserting that Bill ate spoiled meat, there was no such proof produced. When James Clerk Maxwell explained why gases obey the ideal gas law there was.

The ordered pair view developed for explanations of the former kind can be extended to proof-explanations. The product of a proof-explanation will be an ordered triple consisting of a proposition, a proof, and a type of explaining act. The proposition will be a complete answer to *Q*. I shall also suppose that it is one expressible by a sentence of the form "*x* because *y*" in which *x* and *y* are sentences. If the constituent proposition of a proof-explanation is expressible by a sentence of the form "*x* because *y*" then the constituent proof of such an explanation contains proposition *x* as its conclusion and proposition *y* as one of its premisses.

APPENDIX B

A GENERALIZATION OF THE ORDERED PAIR VIEW

Expressions of the form "the explanation of *q* given by *S*" have been said to denote something only if *S* has explained *q*. But we can also refer to explanations that are not products of any acts of explaining—explanations no one has given—

using expressions of the forms "the explanation of *q*" and "the explanation of *q* that is ϕ ." The use of the term "product-expression" will now be extended to cover expressions of these forms which may denote entities even if no explaining acts have been performed. We might say that they denote entities that if not actually products of explaining acts are so potentially. Can the ordered pair view provide denotation conditions for this more general class of product-expressions?

The only difference between "the explanation of *q*" and "the explanation of *q* given by *S*" is that the use of the latter but not the former entails that there was some act in which *S* explained *q*. This entailment is reflected in (iii) of the ordered pair denotation condition for "the explanation of *q* given by *S*." If (iii) is dropped and (iv) is suitably generalized so as not to refer to a particular explainer, we obtain

Denotation condition: "the explanation of *q*" denotes (*x*, *y*) if and only if (i) *x* is a proposition that is a complete answer to *Q*; (ii) *y* = explaining *q*; (iii) $(\forall a)(\forall S)(\forall u)[a \text{ is an act in which } S \text{ explains } q \text{ by uttering } u \supset (\forall r)(r \text{ is associated with } a \equiv \neg x)]$.²³

(iii) does not require that any particular act of explaining *q* has occurred but only that if any does then the proposition *x* is the only proposition associated with it. To obtain a denotation condition for "the explanation of *q* that is ϕ " (e.g., "the explanation of *q* that is the least plausible") we retain (i) and (ii) above, replace (iii) with

$(\forall a)(\forall S)(\forall u)[(a \text{ is an act in which } S \text{ explains } q \text{ by uttering } u \text{ and } x \text{ is the one and only proposition associated with } a) \supset a \text{ is an act in which } S \text{ explains } q \text{ in the } \phi\text{-manner}]$,

and add

$(\forall a)(\forall S)(a \text{ is an act in which } S \text{ explains } q \text{ in the } \phi\text{-manner} \supset a \text{ is an act for which } x \text{ is the one and only associated proposition})$.

The no-product view of section 11 can be extended by analogous procedures.²⁴

The Johns Hopkins University

Received December 2, 1975

²³ "The explanation of *q*" in which no further clause is added is often used to mean "the correct explanation of ϕ ." In such a case "explains" in condition (iii) can be understood as "correctly explains."

²⁴ I am indebted to Robert Cummins, Dale Gottlieb, and George Wilson for their suggestions. The work was supported by the N.S.F.

II. JUSTICE AND HIRING BY COMPETENCE

ALAN H. GOLDMAN

THE issue to be settled in this paper regards a general rule for hiring or awarding scarce desirable positions in society. In recent political debates on the subject of reverse discrimination or preferential hiring, the principle of hiring by competence has seemed to remain sacrosanct, at least if one can judge by the lipservice paid to it by all sides of the discussion. Proponents of affirmative action go to great lengths to distinguish minority "goals" from quotas. While strict quotas for raising percentages of blacks and women employed by a fixed date, which would result in strong reverse discrimination, are acknowledged to be incompatible with the maintenance of strict competence standards, percentage goals for minorities toward which good faith efforts are made are held to encourage minority hiring while maintaining existing standards. Opponents of the policy on the other hand seem to feel that by demonstrating how academic standards of excellence suffer and most qualified individuals fail to get positions through pressure for reverse discrimination, they thereby show affirmative action programs in universities to be unjust.

But despite apparent unanimity regarding the principle in the context of this public debate, it has recently come under attack in more sophisticated philosophical circles from both the left and the right. Libertarians argue or imply that corporations or organizations with positions to fill can give them to whomever they choose, that society has no right to interfere in this free process. Corporations like individuals have the right to control their legitimately acquired assets and to disburse them to whom they choose, and the right to freely hire is part of this more general right. Egalitarians on the other hand hold the principle of hiring by competence unjust in rewarding initial undeserved advantages and purely native talents. Individuals do not deserve those initial advantages for which they can claim no responsibility, and hiring by

competence alone often rewards just such chance talents and advantageous initial social positions. I will argue here against these attacks.¹ I will be concerned with two central questions: (1) Does society have the right to impose and enforce any rule of hiring against corporations with positions to fill? (2) If the answer to (1) is affirmative, which principle of hiring ought to be adopted from the point of view of justice? I will argue against libertarians in this area that society does have the right and duty to enforce a principle, and against egalitarians that hiring by competence is just, that with several qualifications it is as just as human nature allows, and that even without them it is more just than seemingly equalitarian alternatives.

I. THE LIBERTARIAN POSITION

The first question to be faced here is why one system of hiring can be judged more just than another at all, i.e. why the award of jobs by private corporations as opposed to the award of other benefits by private individuals involves considerations of justice rather than simply questions of right and wrong. There are situations in which individuals or corporations can make wrong, even over-all morally wrong decisions, without treating anyone unjustly or unfairly. To say that principles of hiring are a matter of distributive social justice is to imply that certain individuals acquire distributive *rights* to certain positions, and that to refuse them these positions is to refuse to grant them what is legitimately due them. The libertarian denies that any such rights exist. He argues that just as Mary has the right to marry whom she pleases, so a private corporation with benefits in the form of jobs to award has the right to hire whom it pleases without interference.²

No one acquires a right to marry Mary, and similarly, argues the libertarian, no one acquires a right to a benefit from a private corporation

¹ For libertarian arguments, see Robert Nozick, *Anarchy, State, and Utopia* (New York, 1974), ch. 7, 8; also Judith Thomson, "Preferential Hiring," *Philosophy & Public Affairs*, vol. 2 (1973), pp. 364-384. For the egalitarian position see Thomas Nagel, "Equal Treatment and Compensatory Discrimination," *Philosophy & Public Affairs*, vol. 2 (1973), pp. 348-363; also John Rawls, *A Theory of Justice* (Cambridge, 1971), pp. 75-90.

² The analogy is from Thomson; *op. cit.*

which it has not contracted away. It will be useful in criticizing this position to see how far this analogy can be pressed. The difference between the case of Mary and that of the corporation cannot lie in the fact that a person's vital interests are affected by the job he works at, since his vital interests appear equally affected by his spouse, and yet as we said no one acquires a right to marry Mary—she can choose as capriciously as she wants. Thus we cannot argue simply from the fact that it makes a great deal of difference to people what jobs they get to the conclusion that society has a right to enforce a certain rule of hiring against private corporations. Nor does the converse seem to create a distinction, for if Mary's right to choose derives from the fact that her vital interests are involved as well, the same holds true of the corporation's vital interests in its personnel. Since a corporation's welfare and even continued existence depends upon who occupies its various positions, it can be argued that these choices should be left to it. It seems then that just as Mary has the right to choose a husband who will make her unhappy in the long run, i.e. she treats no one unjustly in doing so even if some other suitor would make a perfect spouse, so a corporation has the right to hire total incompetents if it chooses to act so unwisely. Nor can we argue simply that a corporation has no right to hire whom it pleases since the consequences of such freedom are bad (given present biases against minorities). For we perhaps can think of more scientific (or traditional?) ways to match spouses in comparison to which free choice has had long range consequences for happiness, yet we would not want to deny Mary that right, nor contract it away ourselves. Even gains to her own interest or happiness do not justify interference with Mary's free choice, so that her right to choose derives from more than calculation of interests in particular cases. So in the case of a corporation, while it is difficult to see how hiring the most competent could damage its long range interests, perhaps it has a right to ignore those interests if it so chooses.

Are the cases then really totally analogous in relevant ways? First, although interests seem parallel as seen above, are the basic rights involved indeed similar? In Mary's case the general rights underlying her particular freedom in this case include a right over her own body and the freedom to control her life as she sees fit. That these rights, especially the first, have wide scope and absolute priority within their domains is in the interest of all to recognize. In the case of the corporation, the

rights presumably involved are at most weaker versions or narrower cases of these: namely the right to property and that of free association. A corporation may be said to have a property right in the positions it chooses to fill in virtue of having legitimately acquired the assets with which to fund the positions. A corporation like an individual has a right to control those goods or assets which it has legitimately acquired. And the right to control its own assets is empty unless it is free to disburse them as or to whom it chooses. Since present members of the corporation must associate with new appointees, the freedom to associate with whom one pleases may also be cited in support of the libertarian position here. An enforced rule for hiring may force present members to work closely with others against their will, making their work unpleasant for them. And the friction created by this forced close association may be detrimental to the continued smooth operation of the company or organization.

Clearly the right over one's body, which applies in the case of Mary's marriage, is more basic than the right to external property, which applies in the hiring case (whether or not Locke is right in claiming that the latter can be derived from the former), and the freedom to control one's life broader and more precious in total than that of free association. It can nevertheless be argued that just as the former rights constitute the paramount considerations in Mary's case, so do the latter in the case of the corporation. The corporation's property rights to control the disbursement of its assets and the right of free association of its members can be held to imply a specific right to hire whom it pleases without interference from society. In relation to the analogy with Mary's specific right to choose a spouse following from her rights over her own body and to control her life plan, the central questions here are first, whether Mary's rights always entail that she cannot treat others unjustly in choosing a spouse (i.e. whether these rights are absolute in this sphere), and second, whether there might be other rights involved in the case of the corporation, but not in Mary's case, which limit or override those to which it can appeal in support of free choice. To the extent that Mary's specific right to marry whom she pleases is not absolute, and to the extent that the two cases are disanalogous regarding the rights and interests involved, we cannot argue from Mary's case to an absolute right to hire freely of the corporation.

It should first be noticed regarding the question of the scope of Mary's right that she can treat Dick

unfairly or unjustly despite her right to choose, if she has led him on and then rejected him in favor of another at the altar. There are situations in which someone might acquire a legitimate expectation to marry Mary and be treated unjustly or unfairly by her subsequently. This might lead us to suspect that injustice in hiring as well has to do with thwarting legitimate expectations arising from previous efforts. But what could render expectations of individuals to jobs legitimate, given the corporation members' rights of property and free association? It might be held that the only parallel would be a corporation's refusing a job to an individual promised that position, but there are contract laws to prevent that from occurring, and the libertarian acknowledges the state's duty to enforce contracts freely made. (At least the rights in question, as in Mary's case, are already shown to be somewhat limited in scope.)

But there is a difference from the average case of whom one chooses to marry in that whether corporations have competent people or not affects the goods they produce for the rest of society, while whom one marries affects basically only oneself. If Mary happened to be a seventeenth century queen of England and her marriage affected how the country was ruled, she would lose her right to choose whom she pleased, and her subjects could complain of a choice being unfair to them. Similarly it seems society can complain if it fails to get necessary goods and services because of incompetents in positions of responsibility. If one corporation hires incompetents or relative incompetents because of discriminatory practices, it will soon be driven out of business in a competitive situation, but if there is such a practice generally in a whole sector of the economy, the public can complain for the price it pays for such lack of efficiency. To assume that competition will root out all such practices is to oversimplify motives and knowledge of both producers and consumers, and this assumption has proved empirically false.

It may be asked, however, how a social interest in more material goods and services can override recognized rights of individuals or private corporations within the society, like the rights of property and free association here. To determine a social interest is not necessarily to demonstrate the right of society or the state to further that interest, especially when individual or private corporation rights are apparently ignored in the process. For one principal purpose of recognizing individual rights within a system of social justice is to protect

individuals from losses whenever utilitarian calculations run against them in particular cases. The recognition of the right to property, for example, means that a person will not be dispossessed whenever another is in greater need, although such forced transfer would raise total or average utility in particular cases. Therefore the rights to property and free association, it could be argued by analogy, should not be overridden here by the social interest in maximizing goods and services. A private corporation with assets to disburse for job should be free to hire whom it pleases, even when this results in lower efficiency in its production of goods and services. Efficiency cannot be permitted to override recognized rights, or our rights and freedoms would be fragile indeed. Thus while it may be in the interest of all, even of those in power in corporations, to have the most competent hired there may exist rights to ignore the maximization of interest satisfaction, as in the case of marriage choices.

While I accept the above account of rights as far as it goes, it presents an oversimplified picture when used only in conjunction with appeal to property and free association rights in the context of this libertarian argument for freedom in hiring. What is ignored is the fact that recognition of particular rights, like that of property, is established in the first place in relation to a set of varied social values including welfare, and that such rights are therefore rarely (never?) unlimited in scope, but include exceptive clauses recognizing rights established in relation to other values. My right to dispose of my property as I please does not include a right to dispose of my knife in an editor's chest; my right to use my property according to my own wishes does not include a right to play my stereo at deafening volumes; and my right to spend my assets as I like does not allow me to buy nerve gas, even if I keep it sealed in my basement vault. The above examples represent restrictions of freedom to prevent harm, annoyance or potential harm, but these are not the only possible cases. In the initial formulation of rules and rights the value of freedom may be weighed against those of equality or equity, and welfare, for example. This is compatible with the fact that welfare or utility considerations are no longer applied once the rights have been established and their scopes defined. Thus while it may be in the interest of all to recognize a right to personal property, this right may include an exceptive clause regarding filling jobs by corporations, again in the interest of all. This does not mean that property rights are to be overridden in specific

cases for net gains in social welfare—property would be too precarious in that case—but general exceptive clauses of narrower scope (than any net gain in welfare) are compatible with the existence and protection of specific rights like that of property.

We have not yet won the argument with the libertarian, however. For his position is precisely that freedoms, including those of disbursing property and associating with those of one's choice, may only be limited to prevent harm. And a defender of this position would undoubtedly want to press the distinction here between the interests of individuals in maximizing available goods and services and any potential harm to them from the exercise of these freedoms on the part of corporations in hiring. I am not sure, however, that the distinction between harm and utility can be drawn at all in relation to many positions of responsibility in society, such as pilots, surgeons, police and even automobile, home or toy manufacturers. Relative incompetents in these positions represent not only losses in efficiency, but serious potential harm. Thus the harm principle itself, if it allows prevention of unnecessary risk or potential harm, which it must to be at all plausible, may require enforcement of a rule for hiring the most competent in many positions. On a deeper level it may be questioned whether considerations of freedom can be so sharply differentiated from considerations of equality and welfare when designing basic institutions or establishing rules and recognizing rights. To approach this question in this present context, we must first examine how considerations of equality or equity figure in the issue of hiring by competence, for this is somewhat less obvious than the relevance of social welfare or utility.

There is in fact another right involved in this issue which libertarians ignore—what is generally recognized as the right to equality of opportunity. To allow jobs to be awarded capriciously, especially given deep-seated prejudices known to exist in our society, is to deny equal opportunity for goods in a most blatant fashion. An equal opportunity for jobs is the necessary condition for an equal chance to all other basic goods. Thus the right of equal opportunity, if recognized at all, must also be acknowledged to figure more prominently in the issue of a social rule for hiring than the right to property or free association. The reason for this unequal weight is that the right to property as well as that of free association continue to exist although limited by exceptive clauses regarding corporations' doling out jobs, as they continue to exist with

clauses involving limited redistributive taxation or open housing in the name of equality. But equal opportunity for social goods does not exist at all without equal opportunity for jobs. While redistributive taxation, open housing, and integrated schools are advocated in the name of this right, they amount to little when jobs can be denied to those who have managed to acquire superior qualifications with their help. The enforcement of some rule for hiring stipulating criteria not based purely upon inborn or initial chance factors is the first prerequisite for equality of opportunity, since decent jobs are not only of highest value in themselves, but means to most other valuable things.

This admittedly may not bother the thoroughgoing libertarian. For he will most likely recognize no such right as that to equal opportunity, nor give it any weight at all against the maximization of individual freedoms in regard to property and association. He holds that people have a right (short of harm) to what they have freely and legitimately acquired, and that no general right like that of equal opportunity should be recognized which involves repeated violations of individuals' rights to their acquired property. Have we then finally reached in regard to this issue an impasse in moral argument, uncovered an ultimate clash in moral attitude? Rather than admit this we can plausibly continue the argument by accusing the libertarian first of failing to assume a moral attitude on this issue at all, and second of inconsistency in his appeal to the absolute value of freedom over equality.

We can first point out then that part of what it means to assume a moral attitude is to recognize the moral equality of others (implied in a recognition of their subjectivity, i.e. feelings, points of view, etc.)—to accept rules which could be willed from their positions in the social context, or at least by neutral agents. If this recognition of moral equality or moral community within a system of social rules is to be given content as well as form, it means that the rules must not only apply to all, but as far as possible operate to the good of all. It also means that there is a presumption of equality not only in worth but in material conditions, which must be weighed against other values such as that of freedom in the formulation of more specific moral rules. A minimal moral outcome of this balancing (too minimal for egalitarians) is to formulate rules which result not in equality of goods, but in something approaching an equal chance to acquire goods through effort. Hence the recognition and

protection of this right in social rules, such as a rule for hiring, seem a minimal condition for a moral social system. To bring this argument down to the specific issue at hand, its upshot is that rules protecting equality of opportunity, and specifically the recognition that society has the right to enforce some fair rule for hiring against private corporations, are necessary if distributions of property and other goods are to be just.

I have been speaking thus far as if freedom here is to be balanced against equality and welfare, but also indicated above that absolute liberty with respect to property and association may not result in the overall maximization of freedom desired by the libertarian. For poverty and the lack of satisfaction of basic needs which poverty entails constitute an impediment to freedom as well, to the basic freedom to formulate and pursue a meaningful life plan and control one's life as one desires. This is perhaps the most essential liberty of all, and if it is denied through the operation of a social or economic system which leaves some in need so that others may totally control their property, we can view this as an unwarranted conventional constraint upon liberty (property is only protected in the first place by the social system). It follows that any rule of hiring which results in more goods and services, as long as some of these trickle down to those whose freedom is compromised by want, can be adopted not only in the name of welfare or utility, but to increase freedom as well. This does not mean in general that every increase in welfare is to be counted as an increase in freedom as well, or that despotic states with higher GNP's are to be preferred, but only that no social system can be justified in the name of freedom which leaves those at the bottom constrained within the circle of dire poverty. A society with severe racial biases and no rule for hiring results in that situation. Thus we again arrive at the conclusion that society has the right to impose some rule for hiring against private corporations, this time in the name of freedom. Restrictions upon the freedom of corporations to choose capriciously or invidiously in hiring are necessary to protect or create freedom for those for whom equality of opportunity is its necessary condition.

If rules for hiring are then justified through considerations of utility or welfare and equality of opportunity, why not equal opportunity to marry Mary, or the adoption of mating rules which can be shown to maximize happiness or compatibility in the long run? Do the above arguments apply

equally to this case, and if so mustn't they be dismissed in light of our intuitions against social rules for marriage choices? In answer to the first part of the first question, equal opportunity to marry Mary would amount only to an equal opportunity to win her favor, for that is the only relevant qualification we can presently think of for marrying her (we have no independent reliable criteria for what will make her happy in the long run). We might say that equality of opportunity for passing this purely subjective test exists already (Mary's favor might be won by one who could not have been predicted in terms of knowledge of her prior preferences), or given that Mary has certain relatively fixed prejudices, we might deny the possibility of enforcing any rule (except education against such biases) to create equal opportunity for passing this purely subjective test. In answer to the second part of the question, even if we had independent criteria for happiness in marriage as we do for successful job performance, since whom Mary marries affects herself far more than others, others having at most a peripheral interest, we can leave the choice and its consequences to her. Regarding the consideration of social welfare, the welfare of others is not involved in the average marriage cases as it is in who occupies various productive positions. Regarding the consideration of equal opportunity for those applying for jobs versus marriage consents, equality of opportunity in the latter case is not a necessary condition for equal chances at other goods, hence not a necessary condition for basic freedoms or a just social system overall as it is in the case of jobs. For all these reasons none of my above arguments imply by analogy that society has the right to enforce a rule for mating against Mary.

These last points of difference apply as well to the more important and difficult cases of an individual hiring someone for temporary help, or the small businessman who gives a job to his son. Must we to be consistent apply our rule to all such cases and deny these freedoms as well? The first distinction between the case of the small business or private individual and the large corporation is the interest of the public in their products and services. If a small business is the only source of a vital service or product in a given area, it may be reasonable to demand competents in positions of responsibility. Otherwise it may be unreasonable to demand the proprietor to take the time and bear the cost to advertise the position, etc. (this is especially clear in the case of my hiring someone to unload my rented

35-1086

truck, or similar cases). The right of free association is also more central in the case of a small business, and this was part of the reasoning of Congress in applying nondiscriminatory regulations only to businesses with more than twenty-five employees. Since these differences are real, and since equality of opportunity and social welfare do not require that literally every position in society be open to all, but only a certain proportion of them, we may in applying these rationales establish a rule for hiring only for corporations over a certain size, recognizing that the drawing of a precise line will be somewhat arbitrary.

Thus my argument to the effect that certain individuals acquire rights to certain jobs and that corporations treat them unjustly if they are denied those positions involves two steps; first, society has the right in the name of social utility and equality of opportunity to establish a rule for hiring and enforce it against private as well as public corporations and organizations; second, by satisfying this social rule through effort an individual comes to deserve the position in question. The second step is dependent upon the first, which has been established in this section (the second step will be more prominent in the next). I have now provided criteria for the acceptability of a rule for hiring, i.e. social utility and protection of equal opportunity, without having completely shown how a single rule could meet both. In demonstrating that hiring by competence is the correct rule, we must consider the counterarguments of the egalitarian.

II. THE EGALITARIAN POSITION

If the argument of the last section is correct, that is if a rule for hiring is to be justified in terms of social utility and equality of opportunity, it seems easy to show that hiring by competence qualifies as a just distributive principle. Many major theories of distributive justice, especially liberal theories, agree that practices are to be preferred which result in Pareto improvements, i.e. in benefits to some without loss to others. In regard to social utility, it can be argued that hiring by competence results not only in improvements over alternatives of this sort, but in the production of more goods for everyone in society (strong Pareto improvements). Hiring the most competent analytically entails increased goods and services, for competence is defined in terms of the ability to per-

form in a job by satisfying social demand. (I assume here the ability to judge competence according to qualifications, a difficulty which does not in any case affect the argument that we should *aim at* competence.)

The egalitarian attempts to refute the above justification by arguing that those abilities relevant to awarding jobs on the basis of efficiency, which is the basis for hiring in a pure market economy with only profit motives operating, are irrelevant from the point of view of justice; that in rewarding native talent, intelligence and social position (which even if acquired required an ability for acquisition for which the agent can claim no responsibility), the practice of hiring by competence involves rewards which are arbitrary from a moral point of view. Individuals deserve only those benefits which they have earned. They do not deserve their native advantages and so do not deserve those benefits, including good jobs, which flow from them throughout their lives. A child born rich and intelligent stands a far better chance than do other children of acquiring competence qualifications for desirable positions later in life, yet he cannot be said to deserve that better chance from the point of view of justice, nor thus the job which he eventually gets. Thus increments to social welfare from hiring by competence are a matter of social utility from which questions of justice must be separated. The egalitarian appears to have uncovered a conflict between the two criteria for a just rule advanced in the first section, i.e. maximization of utility through hiring by competence seems inconsistent with equality of opportunity in the present social context, and he claims the moral predominance of considerations of equality over those of social utility.

I do not think that any such radical separation of analyses of justice and efficiency could accord with our intuitions regarding the former, as these are aroused by specific examples. Although we may not all be Utilitarians, it seems we must grant that welfare does at least count as a positive consideration, and certainly at the extreme involved in this issue. Following Nicholas Rescher,³ I would argue that a practice or rule which generates a sum total of goods of 15 units to be distributed in a hypothetical society of 4 individuals in shares of 4, 4, 4, 3 is preferable from the point of view of justice to one which generates 8 units to be distributed in shares of 2, 2, 2, 2, despite its greater inequality. In other

³ Nicholas Rescher, *Distributive Justice* (Indianapolis, 1966), ch. 1, 2, 5.

words, those who could have received 4 units under the first plan could legitimately claim injustice at being reduced so as to equal the lowest share under the second, other things being equal. If a claim of injustice or unfairness at being so reduced in moving from the first socio-economic plan to the second is justified, then aggregate utility in itself must be a consideration of distributive justice, and plans must be *prima facie* preferable which result in larger aggregates. To deny this is perhaps to grant too large a moral force to the feelings of envy and pride. For even the person with the lowest share in our hypothetical case is better off under the first plan than the second, except for the fact that he sees others around him with more. There may seem to be a complication here in that those with less relative income (even though more absolute) may be in a worse position to bid for scarce goods. But these alternatives in relation to choices between hiring by competence or other rules for hiring may be taken to refer to goods available, and not simply income. The argument is that more goods will be available to all if competents occupy productive positions, and this is what the egalitarian wrongly claims to be irrelevant to the choice of a just rule for hiring.

Even if we continue to insist upon a radical separation of justice (in a narrow sense) from efficiency, it seems we must admit that the latter can override the former from a moral point of view when it comes to hiring or filling positions of responsibility. It may be, for example, that those best qualified to be brain surgeons do not most deserve the benefits of those positions from a radically egalitarian view of justice, but surely egalitarians would want them to have the jobs anyway (even if they are not the ones on the operating tables). We pointed out in the last section other jobs as well in which incompetents represent not only losses in goods and services, but potential harm of a serious sort. When we think of all the cases of severe harm such as bodily injury or death which occur from defective products or incompetents conducting vital services, it is clear that lowering ranges of competence further would be something to avoid even if at the expense of equality or fairness in the narrowest of senses to job applicants. Since jobs carry responsibilities as well as benefits, society can legitimately complain if its welfare is sacrificed by a policy which leaves those responsibilities unfulfilled. And victims of avoidable irresponsibility (avoidable through a different system of distributive justice or alternative rule for

hiring) can complain not only of incompetence or inconvenience, but of injustice. Another way of expressing this is to say, as we said in the last section, that the public has a right to be spared such avoidable harm.

There might nevertheless be held to be several assumptions involved in the last paragraphs which can be questioned. An opponent could argue that in holding that hiring by competence alone creates more goods for all I am restricting the scope of the concept of goods to that of material productivity reflected in material wealth, and furthermore assuming that such greater wealth (from following this rule as opposed to alternatives) will filter down to those who end up in less lucrative positions (but might have done relatively better under some alternative rule). Regarding the first assumption, this appears to ignore the dignity or lack of dignity attached to certain positions and presuppose that those denied more prestigious positions because of lack of talent or training would be willing to sacrifice a sense of importance, etc. for material benefits. It may be that those with relatively less competence for desirable positions will nevertheless have more and better material goods available to them overall if there is a general rule for hiring the competent as opposed to alternatives, but they may be willing to sacrifice some of this quantity and quality for more equal chances at prestigious positions and the self-respect they carry. Regarding the assumption that the results of greater productivity will filter down (accepted for the sake of argument in the last sentence), this is sound only for an egalitarian distribution of products throughout society, but the complaint against hiring by competence alone in market economies is not simply that it denies dignity to those less qualified for higher positions, but that it lowers their material income as well. Less desirable positions almost invariably have lower pay scales as well, and hence it is questionable whether those occupying them will have maximum benefits under the rule for hiring the most competent, despite the higher total aggregate of goods produced under that system.

In reply to the first charge, it is not necessary to debate or maintain a restriction of primary or basic goods to material wealth in order to defend my position. That part of the justification of hiring by competence having to do with utility depends not upon considerations regarding those applying for particular jobs, but regarding the rest of society or the public as a whole, who could justly complain

not only of loss of material goods and vital services, but of increased potential harm, if those with less than maximal competence were granted positions with social responsibility. It still might be that relative incompetents would be willing to accept these sacrifices and risks under a different rule, but another essential point is that given the inequality of benefits attached to various positions in the economic structure, raising those with less competence simply lowers those with more, and does not result in any more equalitarian distribution of final products. Those who end up with the lower positions under any alternative rule would certainly be worse off than those incompetent for desirable positions at present, given the same inequality in distribution and a lower aggregate of goods and services. And even with the unequal distribution, there is likely to be enough difference in productivity between systems which award jobs to competents versus those which award incompetence equally, that any gain achieved by other individuals in moving from the former system to the latter would be short lived.

Still, an equalitarian could argue that although moving away from hiring by competence would not in itself result in a more equal distribution of goods, a more random process would at least tend to equalize the chances of all individuals for acquiring the benefits of an unequal distribution. The emphasis can again be placed upon achieving true equality of opportunity. There remains the moral dissatisfaction with a system which rewards native advantages, for we feel at least in the abstract that only those characteristics should be rewarded for which an agent can legitimately claim responsibility. This dissatisfaction is warranted, and it is partly for this reason that ideal theories of justice which legitimize hiring by competence add the qualification that there must be equality of opportunity for acquiring competence. To this qualification regarding education or training can be added another regarding rewards for jobs, for the reward scale for different jobs is at least theoretically independent from the criterion on the basis of which the jobs are awarded. While it is not even the case now that those jobs requiring the longest training or highest intelligence are those with the highest rates of compensation (alas!), the pay scale rather being determined for the most part by social demand and political muscle, there are other factors such as intrinsic agreeableness or disagreeableness of the job which ought to be taken into account. Unlike moving away from the practice of hiring by

competence, the readjustment contemplated here toward equalization of pay scales would perhaps not result in great losses of efficiency, since executives would not in any case resign to become sanitation workers. I have my doubts, however, about many doctors, and in any case social demand would still have to play a significant role for reasons of utility (given present attitudes). But the conclusion of relevance to our immediate topic following from these considerations is that inequalities in benefits and unjust rewards for native differences would be better corrected by adjustments to the pay schedule and within the educational system than by a randomization of the process of hiring for jobs. What is most unjust about the present practice of hiring by competence (aside from the fact that those in power often fail to follow the rule to which they pay lipservice) is the final unequal distribution of benefits which results, and the fact that some have initial headstarts toward those benefits which are insurmountable. But these injustices could be corrected without alterations in the rule for hiring itself or the drastic sacrifices in social welfare that would entail.

Although the above lays the groundwork for later points in this section, I have not yet done justice to the egalitarian argument, for it is exactly at this point that it can be resurrected. Its claims are perhaps not meant to be applied to ideal situations in the educational system and regarding differential pay scales, and in such situations my more radical opponent might even agree that hiring by competence is legitimate. Rather they are intended to apply to a society whose educational system fails to make adequate efforts to correct for initial inequalities and whose system of economic rewards is far from just. Given that those jobs often pay most which also carry the greatest social esteem and are the most agreeable or interesting, given that some, no matter how hard they try, simply do not have the native intelligence, talent or initial economic and social advantage to achieve those positions, and given that the educational system does not even encourage those who might succeed but currently lack motivation, the argument is that any system of hiring attached to this situation which freezes these inequalities is unjust in its distributive results. And if the results of a distributive system are unjust, so is the system. Thus, it can be argued, given the intractability of the unequal and unfair results of the system, it would again be more just to equalize chances for the unequal shares by randomization, for example, rather than hiring by

competence. The point can again be pressed that equal chances to unequal shares are fairer than unequal chances to unequal shares. (A realistic egalitarian would require randomization only for those above a minimal level of competence—he would have to recognize some lower limits of efficiency.)

But the egalitarian still seems to ignore the separability of the pay scales and educational system from the rule for hiring. These are conceptually detachable not simply because the former could be made more just independently from alterations to the rule, but more importantly because the justification of the rule appealed to totally other factors. The questionable steps in his above argument consist in the claims that a system of hiring attached to an unjust schedule of rewards is therefore itself unjust in its results, and therefore also unjust in itself. These premisses would be legitimate in this case only if the justification of hiring by competence relied upon appeal to its final distributive results in terms of relative shares to those hired versus those turned down. But the justification offered above was independent of the justice or injustice of the schedule of rewards attached to different positions, appealing rather to the welfare of the general public in terms of goods and services, to protection against discrimination, and to honoring legitimate expectations derived from previous efforts (to attain competence). Even if the efforts of some are unjustly doomed to failure, or if some are not encouraged to make the effort through neglect, it still seems preferable to reward those who have made efforts and succeeded than to ignore effort entirely. It is true that the final distribution of benefits (jobs plus their rewards) will be unjust if *either* the basis for assigning positions *or* the relative schedule of rewards is unjust, but this does not mean it will not be more unjust if fair principles are violated regarding the former as well as the latter.

That the egalitarian argument fails becomes more clear if we seriously consider its alternative to hiring by competence, for the central question in this section is whether randomization is more just as claimed, given the present educational system and pay scales. Is a job lottery, which equalizes chances for desirable positions, the ideal of distributive justice in this area, given present inequalities as frozen into education and pay scales? The general principle, as we have seen, which does seem sound in general, is that when shares are unalterably unequal, it is fair, other things being equal, to

equalize chances for unequal shares; but of course the crucial issue here, which is generally the case in debates on equality, is whether other things are equal, here regarding job applicants. The central problem for establishing any rule of hiring is the stipulation of criteria regarding which characteristics of individuals are to be considered relevant in awarding the positions. Is winning a lottery to be the only relevant characteristic? It seems at first glance (but I believe at first glance only) the most truly egalitarian.

The central complaint against hiring by competence without sufficient compensatory mechanisms in the educational system was that it tended to reward initial differences for which the agents could claim no responsibility. But certainly a process of random selection in the job market, aside from losses in efficiency, comes out worse on this score. It can be admitted that differences in competence for various positions constitute reasonable barometers of prior efforts to acquire competence only where the educational system has attempted to correct for initial inequalities. But even when such remedial efforts are lacking, the acquisition of competence still represents the expenditure of some effort in a socially desirable way, although the effort is not strictly proportional to the degree of competence acquired. The question is whether it is more just to ignore socially productive effort altogether and make all reward a matter of pure chance, which seems implausible or inconsistent if the only complaint against hiring by competence was that it rewarded chance factors to some degree. The effect of a lottery at any level is to negate differences in previous efforts, and if the cost of negating initial social differences (which can be equalized otherwise) is to render all effort meaningless as a measure of desert as well, it hardly seems worth it from the point of view of distributive justice.

The argument in favor of randomization now becomes that although it awards jobs on the basis of pure chance, at least it equalizes the chances—if benefits are to be doled out according to rolls of the dice, at least the dice should not be loaded. While the dice are loaded at present in favor of the rich, the talented and the intelligent, randomizing the process of hiring restores the balance which should have been restored by the neglectful educational and economic systems. In reply it could be argued first that in the original Book of Life all had equal chances at intelligence, etc., which is just a colourful way of saying that how intelligent you are or who your parents happen to be is a matter of

chance from your point of view: it is simply a matter of chance operating further in the past. It makes a difference whether the chance factor operated in the past or in the present in the form of a job lottery, however, since in the interim have occurred the efforts of those who have attained requisite skills. Second, there is an important difference ignored in the above argument between correcting for inequalities in the educational or tax systems and annihilating them at the final bell in the hiring process. For there is a limit to the justifiable neglect of those with superior talents, motivation and monetary assets, and their efforts should be ignored no more than those less fortunate once reasonable attempts have been made to motivate the latter and make their success possible.

The real contrast still reduces to that between rewarding chance versus rewarding effort and past and potential social contribution, and it still seems that a random lottery for hiring is the worst of all possible worlds by these criteria. Where there is no ulterior social purpose in the reward of some benefit than the distribution of some windfall good, and where furthermore no previous actions of the individuals in question can be seen to create differential rights or deserts to the goods, a random process of choosing is fairest. This follows from the presumption of equality of persons deduced in the first section. But when positions are assigned for socially productive purposes, and when individuals are therefore encouraged to direct their efforts towards fulfilling these purposes, past and potential productivity achieved through these efforts cannot be justly ignored. The same points apply to a lesser extent to randomization for those above a certain minimal level of competence, and in any case the egalitarian's reasoning must lead him to suggest this only for reasons of efficiency and not justice, which he takes to be distinct. Total randomization functions as the ideal of egalitarian justice in this area (given that jobs and their rewards must be unequal), and it is therefore relevant to argue as I have that it is a misconceived ideal.

The argument regarding socially productive effort can lead us to question further whether the supposed egalitarian ideal really represents a net increase in equality. It can first be maintained that randomization in any case detracts from liberty—the liberty to pursue the career of one's choice. It is assumed by proponents that a lottery would increase this liberty for many, since it would allow all who wish to pursue a given position to do so merely by entering the lottery. But the meaningful

pursuit of what one considers valuable and worth achieving involves effort and even struggle, and part of the enjoyment of reward is the feeling of success after effort. Such real pursuit is rendered meaningless by being ignored in awarding positions by chance. The same argument can be put in terms of equality and the denial of equal right—here equal right to pursue value and acquire through effort what one considers worthwhile, as well as or perhaps equivalent to equal right to dignity. Dignity, or a sense of personal worth, beyond the mere dignity of being human often mentioned in equalitarian arguments, derives from a sense of having accomplished, not won by chance. But an equal right to jobs at all levels is bought at the expense of an equal chance to meaningful achievement through effort (certainly in the early and formative years).

It might be held that a randomization of rewards and recognitions does give an equal right to this type of pursuit—namely no right at all. This might seem a pyrrhic victory for the egalitarian, but might at least show that my argument here cannot be stated in terms of equality, but only in terms of balancing other values against that of equality. But when some have directed the efforts of their early years to attaining skills needed for some type of position they hope to occupy, and then see these positions filled through a chance process by others, who acquire at least a small chance of hanging on to them without having achieved the same level of competence through the same degree of effort, there is an inequality and unfairness involved. If to avoid this unfair waste lotteries are pushed further back into the early years, so that individuals know the positions they are to occupy from the beginning and so do not make efforts in the wrong direction, the only relevant differences between this determination and the operation of original chance differences against which the egalitarian complains would be that this assignment of positions would be more rigid and would involve losses in social utility, as compared with more open and widespread competition for productive positions.

The egalitarian could further argue that in his system a sense of dignity or achievement can be developed through performing well on the job rather than in the process of acquiring it, and that his systems equalizes chances for the former type of achievement. Aside from the fact, as mentioned above, that this downgrades accomplishments culminating the first twenty-five years or so of a per-

son's life, the dignity of even those who acquire positions through a random process would most likely not be appreciably enhanced (could they be expected to perform well, having never found the need to do so before while awaiting the lottery?), and the prestige of the positions themselves and the self-respect that goes with them would soon diminish with the knowledge that they were gotten by chance. If anyone could enter any lottery for any position, then no one could pursue positions or feel any satisfaction in having achieved them. If competence were nevertheless later demanded on the job, and a sense of satisfaction were supposed to derive from having performed well at that level in demanding positions, there would result a series of firings and hirings until competent people filled the positions anyway, with a net increase in resentment for those who proved and could have been predicted to be incompetent. If not everyone can enter all lotteries, but only those who have achieved a certain level of competence through effort, and a sense of achievement is to be derived from having acquired that level, why ignore all achievements beyond that level? (If a number of individuals applying are at the same level, there will have to be some other means of choosing anyway, possibly preferably random.) While we do not want to reward purely native characteristics, and to do so is generally unfair, it is an empirical fact which makes my argument somewhat easier that there are few desirable positions for which successful performance demands only native characteristics.

These last arguments really rely upon the abstractness of the notions of liberty and equality, to which other writers have pointed as well, upon the fact that "*Liberty*" and "*Equality*" are generally emotive laden shorthand for "liberty to do *x*" and "equality with respect to *y*." With regard to the latter, the above points illustrate that the crucial question is always what factors should be given equal weight in any decision or distribution, that giving certain factors equal weight often militates against considering others. What I have been suggesting is that equality of opportunity should be defined not in terms of equal or random chances at goods including jobs, but in terms of real and equal chances to achieve success through effort. Proponents of randomization would presumably base their stand regarding the ambiguity of "equality" upon the claim that the most equalitarian and least meritocratic criterion is to consider people equal just in virtue of being human wherever possible, rather than more narrowly specifying the criteria

for the distributions of goods, and that this is the purpose of the job lottery. I argued in the first section that this broad criterion of humanity is relevant in generating a right to have basic needs satisfied and a presumption of equality of conditions from which deviations must be justified. This much is implied in the recognition of the common humanity or distinct subjectivity of others who share a moral community. But unequal efforts and social productivity do seem to justify differences in benefits, and it is doubtful that the egalitarian criterion can be applied at all in the case of job hiring. It has to be admitted right off that a job lottery attempts to consider people equal not simply in virtue of being human, but in virtue of want (desire for a particular position), or in virtue of having entered the lottery, since those who do not want the position enough to enter the lottery will not be considered. But then it becomes questionable whether equal weight is being given to want or degrees of want, since all degrees over the minimal necessary to enter the lottery is discounted. It seems that a better or fairer way to measure want would be in terms of wanting enough to have made the effort to acquire competence, and we are brought back to the argument on rewarding socially productive effort rather than luck.

The final point to be raised regarding effort and its just rewards is that even if hiring by competence is not most just in relation to abstract moral principles, the fact that society regards it as just or has adopted it as a working rule legitimizes certain expectations arising from successful efforts to satisfy that rule. Although I maintained in the first section that an individual's acquiring a right to a position is dependent upon society's right to establish a rule for filling that position, it is somewhat independent of which rule society adopts. Individuals living in a community must for the most part direct their efforts toward socially acceptable goals, and if these efforts are constantly thwarted after the fact by changes in social criteria for success, the psychological effect is devastating, and society is to be held responsible. While this conservative line of reasoning is abused or overstated if used to apologize gross injustice in the status quo, it does create a presumption in favor of honoring prior commitments of the community in relation to its accepted rules, just as there is a presumption which can nevertheless be overridden in favor of keeping promises, even when what is promised is not independently morally best. Thus we can speak of rights in practice or rights created through satis-

III. DOING, REFRAINING, AND THE STRENUOUSNESS OF MORALITY

CAROLYN R. MORILLO

A GOOD deal of current ethical debate has centered on a cluster of closely related distinctions, stated in more or less general form and with more or less general implications for one's ethical theory. We have been asked to compare killing and letting die, avoiding producing harm and preventing harm, negative and positive duties, doing and refraining, positive and negative responsibility, commissions and omissions, and so on. The key initial question has been whether any of the above distinctions locate differences which are morally relevant to the rightness or wrongness of actions and the responsibility of agents. Is there more, morally, to be said against killing than against letting die? Am I as responsible for the harm I fail to prevent as the harm I directly inflict? What do the distinctions tell us about hard cases where we must choose among evils, and particularly those where we would have to harm some in order to save others?

It is a reasonably accurate generalization to say that those who have argued for there being morally relevant differences in these distinctions have had an affinity for absolutist, "conservative," or just non-consequentialist views, while those who have denied that the differences are morally relevant have held consequentialist and, in some sense, radical views. There are obvious exceptions to this (Philippa Foot defends the moral distinctions between negative and positive duties, but could scarcely be called an "absolutist"), and the truth of the generalization may partly be due to the historical accident of so much of the debate focusing on the abortion issue. At the same time, it has, for example, been argued by Bernard Williams that "consequentialism essentially involves the notion of *negative responsibility*: that if I am ever responsible for anything, then I must be just as much re-

sponsible for things that I allow or fail to prevent, as I am for things that I myself, in the more everyday restricted sense, bring about."¹ And one of the strongest defenders of negative responsibility, John Harris, argues that there must be a "radical revision of our views about responsibility," and he draws from his doctrine some startling moral implications.² Again, Peter Singer argues that "if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it."³ The implications he draws from this are that "the traditional distinction between duty and charity cannot be drawn, or at least not in the place we normally draw it,"⁴ and more generally that "the whole way we look at moral issues—our moral conceptual scheme—needs to be altered, and with it, the way of life that has come to be taken for granted in our society."⁵

I think there are vital issues here, though not perhaps always where the disputants have located them. I would like, in this paper, to work at three related tasks. First, I shall try to work out a general topography of the above distinctions, which often seem to cut in such various ways. My aim will be to provide a single, coherent (though definitely not non-controversial) classification of the distinctions which will help focus the moral issues. Second, I shall examine a number of the arguments which have been offered to account for and defend the moral relevance of the distinction. I shall, in general, be drawing the conclusion that the arguments do not work, but in specific instances I shall try to do justice to certain residual moral concerns which I believe are expressed in the arguments. All of this will, in a rather negative way, place me in the camp of the radical consequentialists. At the same time, I shall offer

¹ J. J. C. Smart and Bernard Williams, *Utilitarianism For and Against* (Toronto, 1973), p. 95.

² "The Marxist Conception of Violence," *Philosophy and Public Affairs*, vol. 3 (1974), p. 198.

³ Cf., "The Survival Lottery," *Philosophy*, vol. 50 (1975), pp. 81-87.

⁴ "Famine, Affluence, and Morality," *Philosophy and Public Affairs*, vol. 1 (1972), p. 231.

⁵ *Ibid.*, p. 235.

⁶ *Ibid.*, p. 230.



examples of some of the facts of our times which give positive reasons for accepting such a position. Finally, I shall argue that these very same facts, combined with my suggested schema for classifying actions, show that for a large range of absolutely crucial issues there is today not even a "pragmatic" line between positive and negative responsibility, and that the simple determination not to harm others must involve us in very radical changes indeed.

I

Since the disputes with which we are concerned are ethical, any suggestion as to how we should describe the actions being assessed is inescapably controversial. (It might seem at this stage question begging to use the generic term "action" to cover doings and refrainings; if so, think of them collectively as act-like situations.) Shall we say that if someone, *A*, deliberately and knowingly lets someone else, *B*, die, that *A* has killed *B*? That is not a question we can settle prior to dealing with the moral issues. Furthermore, anything like a complete discussion would have to develop a complete theory of action.

At the same time, there are clearly certain dimensions of and perspectives on these actions which all would acknowledge as morally relevant. And what we can require is that the descriptions of the actions—if not definitive—at least be appropriately symmetrical. We are being asked to compare the moral import of actions. As a minimum condition, then, one must use the same principles of description in characterizing those actions. As I shall hope to show later, simple attention to this requirement is sufficient to undercut a number of the moral arguments in the literature.

I shall introduce my classifications by considering the much discussed example—killing versus letting die. What is it about these possible situations which concerns us morally? What do we want to know in order to be able to make a reasonable judgment?

First, there is the fact that a certain state of affairs has occurred, namely, the death of some living creature (let us assume it is an innocent human being). Change the nature of that state of affairs and you clearly alter the moral significance of the cases. (Compare dropping the shoe on the floor versus letting it remain on the floor; it would be the unusual case in which we would be concerned about the state of affairs of a shoe being on the floor.)

Second, we want to know what the connection is between some agent and this state of affairs. How does killing relate the agent to that result as compared to letting die?

Finally, we want to know about the beliefs, intentions, and motives of the agent, as well as various costs to the agent.

In addition to these dimensions, it can make a difference whether we describe our cases from the perspective of a spectator with the knowledge of what actually occurred, or from the perspective of the agent making a decision about what to do. Suppose *A*, a lover of all things strange and reptilian, holds out toward *B* a harmless (i.e., non-venomous) snake, in order to show the beautiful patterns on the skin. Unbeknown to *A*, *B* has a pathological fear of snakes and a very bad heart condition. As a result, when the snake is brought toward *B* he collapses and dies. Given the knowledge of this death, the question of the probability of the death is either moot or we should have to say that the probability of it has become one. But from *A*'s perspective, we certainly would not say that he could or should have judged the probabilities that way. There has, in a sense, been a killing, though the moral pressure of *A*'s perspective would usually lead us to feel that the "fairer" description would be "accidentally caused the death of *B*." By contrast, suppose *C* deliberately shoots *D* in the head three times. With extraordinary surgery, *D* survives. Given that information, the probability of *D*'s death is zero, but that would certainly not have been the reasonable probability judgment for *C* to have made. There has been no killing, but *C* may well be charged with attempted murder.

Let us label these two perspective the Results Perspective and the Decision Perspective. I shall now sketch a way of characterizing our act-type situations which is applicable to either perspective and which provides, I believe, a morally illuminating way of viewing the connections between agents and states of affairs.

I shall start with these informal assertions. As agents, we implement decisions. At least one question we must ask, as rational moral agents, is, What are the probable consequences of my implementing this decision as opposed to alternative decisions I might implement?

These assertions suggest that we may view our act-like situations as involving the following components:

(1) *An implemented decision (ID)*. I shall use this phrase to cover the carrying out of a decision to

perform or not to perform certain bodily movements. It is not to include any events or states of affairs beyond the agent's own body, no matter how those are related or are believed to be related to his bodily performance or non-performance. This will entail in many instances a somewhat artificial redescription of the agent's decision. For example, suppose *A* aims a rifle at *B* and then decides to kill *B* and so pulls the trigger. My suggestion is that we say *A* implements the decision to move his trigger finger in the appropriate way, thus leaving it an open question whether he has implemented the decision to kill. The latter implementation would depend, among other things, on the gun firing, *B* dying, and so on.

(2) *An ethically significant state of affairs (ESS)*. This will be any state of affairs which has positive or negative value, at least in the sense that if there is any probability that the state of affairs will result from an implemented decision, then a rational moral agent would have to take this into account as being morally relevant to his decision. In the killing/letting die cases, that state of affairs is the death of some living creature (perhaps a human). For the purposes of this paper we do not need to work out a full theory of value. All that is required is that one be consistent in the valuation placed on a given state of affairs.

(3) *Probability connections* between implemented decisions and the ethically significant states of affairs. These may be characterized from either of the two perspectives mentioned above.

(a) *Decision Perspective*. In a full classification, this would include any actual probability beliefs an agent might have, no matter how ill founded. For our purposes, however, we may restrict our consideration to the reasonable probability beliefs in the given situation. Given a rational agent who knows what he could reasonably be expected to know, what would it be reasonable for him to believe would be the probability of the ESS if he implements a given decision, as compared to its probability if he does not.

(b) *Results Perspective*. What in fact was the effect of the ID on the probability of the ESS, as compared to what the probability of it would have been in the absence of the ID. Our initial discussion will focus on the two extreme cases, where either the ID

makes the probability of the ESS zero or the ID makes the probability of the ESS one. It is important to note, however, that we can allow for all degrees between these two extremes, which should be helpful in dealing with statistically complex phenomena with a large number of contributory causes.

I need now only to add a few additional refinements, with appropriate abbreviations, in order to have the machinery for setting out a preliminary classification of the act-type situations with which we are concerned.

Implemented decisions to perform a certain bodily movement I shall abbreviate as ID_D (as in doing). Implemented decisions not to perform a certain bodily movement I shall abbreviate as ID_R (as in refraining). It should be noted that one who is in this sense "refraining" may well be engaged in bodily movements other than the one which the agent has decided not to perform. Refraining does not entail the absence of all movement, unless the ID_R is to refrain from *all* movement.⁷

Those ethically significant states of affairs which have positive value will be abbreviated as FESS, while those with negative value will be NESS. Once again, no specific theory of value is intended beyond the implication that if an ID is likely to produce a PESS this tells for implementing that decision, while if it is likely to produce a NESS this tells against the decision.

The probability connection between an ID and some ESS will be symbolized as follows. If the ID in fact makes the probability of the ESS one then this will be written $ID \rightarrow (1p)ESS$. If the ID in fact makes the probability of the ESS zero this will be written $ID \rightarrow (Op)ESS$. This symbolism can be expanded to accommodate degrees of probability, rough classifications such as high or low probability, or even the notion of raising or lowering the probability, but this will not be necessary at this stage. All of this is stated from the Results Perspective. It will be possible to indicate informally when we are dealing with the Decision Perspective, that is with a reasonable belief about the effects of the ID on the probability of the ESS, rather than with the actual effect. (As our knowledge increases, the probability estimate of the Decision Perspective will tend to approximate that of the Results Perspective.)

We may now list the following combinations.

⁷ I would like to thank the referees for this paper for calling my attention to the need to clarify this point, as well as for other helpful comments.

- | | |
|--------------------------------|--------------------------------|
| 1. $ID_D \rightarrow (ip)PESS$ | 5. $ID_D \rightarrow (ip)NESS$ |
| 2. $ID_R \rightarrow (ip)PESS$ | 6. $ID_R \rightarrow (ip)NESS$ |
| 3. $ID_D \rightarrow (Op)PESS$ | 7. $ID_D \rightarrow (Op)NESS$ |
| 4. $ID_R \rightarrow (Op)PESS$ | 8. $ID_R \rightarrow (Op)NESS$ |

Let us suppose that in each case the knowledge of the agent is such that his reasonable belief about the probability of the ESS approximates the actual probability (i.e., he is "morally certain" of the outcome of his implemented decision).

We may now consider what informal labels are appropriate for each of the above cases. This will immediately make clear—if it is not already—why this method of classification is controversial. Looking at the table, does it not seem quite appropriate to label both 1 and 2 as cases of bringing about positive value (e.g., benefit), 3 and 4 as preventing benefit, 5 and 6 as bringing about negative value (e.g., harm), and 7 and 8 as preventing harm. In each case the implemented decision *has as a result* the occurrence or non-occurrence of some ESS, in the sense that if that decision had not been implemented there would not have been such an occurrence or non-occurrence (i.e., the probability of the ESS would not have been one or would not have been zero). What the table stresses is the symmetry of these pairs, and this, of course, is the issue. To take just one example, counting the death of an innocent human as negative, killing would be a case of 5 and letting die of 6. Again, actively preventing harm—a "positive" duty—would be 7 while refraining from producing harm—a "negative" duty—would be 8.

We have, then, two questions. The first is whether there are any morally significant differences between the above pairs, 1 and 2, 3 and 4, 5 and 6, 7 and 8, and if so, what they conceivably could be. Presuming motive, intention, knowledge, and cost to the agent to be held constant, the only difference seems to be the presence or absence of some particular bodily movement. With respect to that I feel inclined to say that for rational, decision-making creatures, the mere presence or absence of such a movement is simply *not* morally relevant—that what matters is the decision we implement and how this affects the probability of some ethically significant state of affairs.

The second question, however, is whether the contrasts with which people have been concerned are really captured by the above table or by any extension of my scheme of classification. Is there something essential to—definitive of—doing versus

refraining, or negative versus positive responsibility, which I have missed and which could account for morally relevant differences?

To take just one example (which may already have occurred to the reader), in comparing 5 and 6 in our table, if an implemented decision to move in a certain way results in a probability of one for some negative state of affairs, the assumption is that in the absence of that movement the probability would not have been one, indeed would have been zero (i.e., the NESS would not have occurred), and this surely is a case of bringing about the NESS. By contrast, if the implemented decision is *not to move* in that way then the assumption is that the presence of the movement would have made the probability less than one (perhaps zero), but in addition there is the assumption that the situation prior to the agent's decision is already such as to make the probability of the NESS one unless the agent moves in the appropriate way. If the agent decides not to move, then he has indeed failed to prevent a harm which was, so to speak, underway, but is that the same as bringing it about? There are important points here, having to do with levels of responsibility and with placing at risk, but I shall argue below that these are among the "residual" moral issues with which I said I would deal, and that they *do not* alter the moral symmetry between 5 and 6, or any of our pairs.

There is no short way with these questions—we must examine cases and actual arguments. I would stress again, however, how absolutely essential it is that any principles of description be applied symmetrically. Additional variables must be held constant. If one take a Results Perspective in one case, one must use the same perspective in describing a case with which it is to be compared. With that warning, we can proceed.

II

Symmetry of description requires that both probabilities and perspectives be held constant. I shall now examine several arguments for the moral relevance of our distinctions which fail because they do not meet this requirement. The failure is particularly surprising in that case where the argument was written as a reply to Jonathan Bennett's well known article, "Whatever the Consequences."⁸ In that article, Bennett is arguing against the moral relevance of the acting/refraining distinctions underlying what he calls the "conservative"

⁸ *Analysis*, vol. 26 (1966); reprinted in Thomas Beauchamp, *Ethics and Public Policy* (Englewood Cliffs, 1975), pp. 328–348.

position on the taking of one life in order to save another. He quite clearly makes the points which I shall stress in this section, but judging by the responses the points have still been missed.

Bennett notes that in many cases involving a choice between doing and refraining (as in the obstetrical cases on which he focuses) there are morally relevant differences, but that these differences depend on factors which would be acknowledged by any consequentialist and which do not depend on the doing/refraining distinction *as such*. Among these would be probability factors. Bennett states that "moral conclusions may be supported by facts (a) about what is expected . . . ; (b) about what is inevitable. . . ." ⁹ He explains these factors in terms of what he calls the "upshots" (in his cases, the death of the mother or of the child) of our performing or not performing certain physical movements. These clearly parallel those ethically significant states of affairs external to the agent which I have used in my schema. He characterizes factors (a) and (b) as follows: "(a) . . . whether in making the movement I was entirely confident that that upshot would ensue . . . (b) . . . how certain or inevitable it was that that upshot would ensue from my movement, or upon how confidently I ought to have expected it to ensue . . ." ¹⁰ Factor (a) seems to cover any probability belief the agent may have, whether reasonable or not. Clearly it involves what I have called the Decision Perspective. It is not so clear how to handle (b). In so far as it stresses what the agent's expectations *ought* to be, it seems to be dealing with reasonable probability beliefs from the Decision Perspective. At the same time, the phrase "how certain or inevitable it was" seems to be pointing toward actual connections between movements and upshots, not just reasonable beliefs about such connections, and to that degree is closer to my Results Perspective. For our purposes, these discrepancies are not critical. Most of Bennett's cases are ones in which there is "moral certainty" about the relevant upshots.

The central point is this. Bennett clearly shows that if we are to demonstrate a morally relevant difference between killing and letting die, or between acting and refraining *as such*, we must factor out (i.e., hold constant) the probability

connections between doings or refrainings and upshots (between my IDs and ESSs). If, in addition, we factor out knowledge, intention, and all other differences not definitive of the doing/refraining distinction, then, Bennett concludes, any remaining differences are morally irrelevant. ¹¹

Let us now consider a criticism of Bennett by Daniel Dinello. ¹² Dinello offers the following analyses of "killing" and "letting die," and claims that the distinction as he draws it does have "some moral bite."

- (A) x killed y if x caused y 's death by performing movements which affect y 's body such that y dies as a result of these movements.
- (B) x let y die if
 - (a) there are conditions affecting y , such that if they are not altered, y will die.
 - (b) x has reason to believe that the performance of certain movements will alter conditions affecting y , such that y will not die.
 - (c) x is in a position to perform such movements.
 - (d) x fails to perform these movements. ¹³

In this analysis, " x killed y " is clearly characterized from the Results Perspective, whereas " x let y die" is, at least in condition (b), characterized from the Decision Perspective. This immediately weighs the cases. In the first place, any causing of death, no matter how accidental or unintended, is to be labeled a "killing." More to the point is the fact that on the killing side we are led to think of the death as certain (it did happen), whereas on the letting die side we are to focus on the fact that x only has "reason to believe" that he could prevent y 's death, not that it was certain that he could. For a fair and symmetrical comparison either (A) should be rewritten from the perspective of the reasonable belief of x , or (B) (b) should be rewritten to state simply that the performance of certain movements by x would alter the conditions affecting y such that y would not die. The latter change would not, of course, be plausible, since we would not want to say that x let y die if there were movements x could have performed which would have prevented y 's death, but x had *no way of knowing this*. But in the same way, if x performs movements which in fact cause y 's death, but x had *no way of knowing this* would be the result of his movement, we would not want to say that x killed y .

⁹ *Ibid.*, p. 333.

¹⁰ *Ibid.*, p. 333.

¹¹ *Ibid.*, p. 341.

¹² "On Killing and Letting Die," *Analysis*, vol. 31 (1971), reprinted in Beauchamp, *op. cit.*, pp. 355-358.

¹³ *Ibid.*, p. 357.

One point is that for any issue of responsibility or moral judgment to arise, some knowledge or reasonable belief condition must be added. At a minimum, we would want to be able to say of x that he should have known or should have believed that the relevant ESS (in this case, y 's death) would result from his movements, or from his failure to move (in the sense that he would have prevented the ESS if he had performed certain movements). We must assume the same actual probability connections and the same beliefs about or knowledge of those connections on the part of the agent before we can make a fair moral comparison. This is why I suggested above that we presume full knowledge ("moral certainty") on the part of the agent implementing the decision, in order that we might concentrate on the actual connections between doings or refrainings and some ESS.

In fairness to Dinello, he does attempt to hold all such factors constant in an example he offers which he believes supports his conclusion that "it seems intuitively clear that causing a death is morally somewhat more reprehensible than knowingly refraining from altering conditions which are causing a death."¹⁴ His case is as follows:

Jones and Smith are in hospital. Jones cannot live longer than two hours unless he gets a heart transplant. Smith, who had had one kidney removed, is dying of an infection in the other kidney. If he does not get a kidney transplant, he will die in about four hours. When Jones dies, his one good kidney can be transplanted to Smith, or Smith could be killed and his heart transplanted to Jones. Circumstances are such that there are no other hearts or kidneys available within the time necessary to save either one. Further, the consequences of either alternative are approximately equivalent, that is, heart transplants have been perfected, both have a wife and no children, etc.¹⁵

Dinello states that "it seems clear that it would, in fact, be wrong to kill Smith and save Jones, rather than letting Jones die and saving Smith." In fact, he argues that even if Jones has a wife and Smith does not, it would still be wrong. Thus he concludes, "There are cases where consequences are equivalent and cases where consequences of killing are preferable, yet still wrong to kill."¹⁶

¹⁴ *Ibid.*, p. 358.

¹⁵ *Ibid.*, p. 357.

¹⁶ *Ibid.*, p. 358.

¹⁷ The issues here are complex. For a more detailed discussion of a case similar to Dinello's—though it involves killing one in order to save two—see John Harris, "The Survival Lottery," *op. cit.* In my comment on Harris, "As Sure As Shooting" (*Philosophy*, vol. 51 [1976], pp. 80–89), I try to point out the contemporary relevance of Harris' case as well as some of the wider issues involved.

¹⁸ "Saving Life and Taking Life," *The Journal of Philosophy*, vol. 72 (1975), pp. 131–137.

Hovering in the background here are all sorts of wider issues concerning medical policy, the weighting of lives, and so on. But presuming all of this is held constant, then the moral heart of this case is the same as that in Bennett's more realistic obstetrical case—we may kill A to save B , or else let B die in order to save A . Well, what should we say in such cases? I think here is one point where consequentialism would—is intended—to radicalize our intuitions. We should not misinterpret, however, just what this involves. The point of the radical criticism is not to find fault with what our intuitions have been telling us about killing, but with what they have *failed to tell us about letting die*. In the Dinello case, what we would like to think is that Jones' death is inevitable (because *earlier* than Smith's?), that he is going to die "anyway," and that we can then go about the business of saving Smith. But this is not true. As the circumstances have been described, we *could* save Jones. Or we could save Smith. But not both—which is the tragedy of the choice. What we cannot escape is that with *either* choice we are sacrificing a human life—that letting Jones die when we could save him is as much a sacrifice as killing Smith in order to save Jones.¹⁷

There are other moral issues here, which I shall discuss later, concerning responsibility and placing at risk, but they will not alter the conclusion that any moral distinctions to be made cannot be grounded on the distinction between killing and letting die.

There is even more explicit use of probability factors in an article by Richard Trammell.¹⁸ Showing how he misuses these factors will also clear up some ambiguities in such notions as "letting die" or "failing to save." Trammell is discussing Michael Tooley (his "Abortion and Infanticide") instead of Bennett, but the basic issues are the same.

The focus is on "negative" versus "positive duties," the duty to refrain from causing harm versus actively preventing harm. These are 7 and 8 on my table rather than 5 and 6. Trammell is arguing for the view that "in some cases we are under greater obligation to avoid taking a life

than to save a life, even though effort and motivation are constants,"¹⁹ that negative duties are more stringent than positive. The first thing of interest to note is that in this case greater moral weight is placed on the *refraining* rather than the *doing*, which is just the reverse of the killing/letting die cases (my 5 and 6) where, it has often been argued, the *doing* is more serious than the *refraining*. This in itself should make us suspicious about any moral distinctions being placed on doing and refraining as such.

One of the factors which, according to Trammell, underlies the distinction between negative and positive duties is what he calls "optionality." As he explains it, "Some actions either destroy a good or make it impossible for anyone else to realize a certain good; whereas other actions do not destroy and perhaps leave open to others the option of realizing the good in question." He then argues as follows:

... there is an essential difference between negative and positive duties in regard to the probability of the good being realized in case the duty is not met. There is a logical equivalence between *y* failing to meet the negative duty of not causing *x* to suffer at time *z*, and a probability of 1 that *x* suffer at time *z*; whereas there is no logical equivalence between *y* failing to meet the positive duty of relieving *x* from suffering at time *z*, and a probability of 1 that *x* suffer at time *z*. If *x* kills *y*, it is certain that *y* will not live; but if *x* fails to save *y*, someone else may still have the option of saving *y*.²⁰

It should be evident what the error is here. Trammell is applying a Results Perspective to negative duties, but a Decision Perspective to positive duties. In addition, he fails to note a critical ambiguity in the concept of "failing to save." It is true, for example, that if *x* kills *y* there is a probability of 1 that *y* dies. Nothing *x* did could be called a killing unless someone actually dies. This is part of the logic of "killing." What, then, about "letting die"? Here too we would not say of *x* that he let *y* die unless *y* actually dies. That is, there is a logical equivalence between "letting *y* die" and a probability of 1 that *y* suffers death. This is not so clear, however, if we say "*x* left *y* to die." We might say this even though *y* is rescued. The ambiguity is certainly there in "*x* failed to save *y*." This could be said (perhaps with the emphasis, "*You failed to save him*") with the

implication that *y* was *not* saved, but it could also be said ("*You failed to save y*") in cases where someone else did save *y*.

The point is this. If we are to compare negative and positive duties, then we must compare cases where the probabilities—actual or believed—are the same. If we start with killing, then its analogue must be letting die. On the other hand, if our "failing to save" case is one where it is reasonable for *x* to believe that there is some probability that *y* will be saved anyway, then the moral analogue is a case where *x* performs a movement which puts *y*'s life at risk but where it is also reasonable for *x* to believe that there is some probability that *y* will not die. If the probabilities are the same, is it so clear that there is any moral difference?

Trammell says that "we feel obligated to go to almost any length to avoid killing someone, but not under equally great obligation to save someone."²¹ If two situations were as follows—(1) if I perform certain movements then it is certain, and I know it is certain, that *y* will die, and (2) if I refrain from performing certain movements it is certain, and I know it is, that *y* will die—then would we feel a greater obligation to refrain from moving in (1) than to move in (2)? (It is assumed, of course, that effort—"cost" to the agent—is the same in each case.) If so, then I would say we have made a moral mistake.

On the other hand, if, more realistically, we focus on all those cases where we are failing to save but where there is only some probability (and considerable uncertainty as to just what it is) that those we are not saving will die, then we should equally focus on all those cases where we engage in movements which have a comparable probability of causing death. Once we are in the arena of probabilities and statistics (the main arena for all of us today) the range of cases is enormous. Any environmental activist could compile his own cautionary list—driving at 70 m.p.h., using aerosol sprays, using pesticides, eating tuna-fish (presuming, as I do, that the deaths of hundreds of thousands of porpoises in the tuna fishing nets is an ethically significant state of affairs), polluting a river, and so on. As we learn more and more, statistically, about the deaths that lie at the far end of so many of our activities, is it true that we are prepared to go to "almost any length" to avoid these? Not, apparently, if we would have to give up

¹⁹ *Ibid.*, p. 133.

²⁰ *Ibid.*, p. 135.

²¹ *Ibid.*, p. 133.

some amenities, or if the dollar cost seems too high.

To summarize my arguments so far, much of the presumed moral difference between killing and letting die, or negative and positive duties, or any of the pairs on my list would seem to be based on differences in probabilities, a purely consequentialist consideration. If we are careful to ensure symmetry of probabilities, there then seems a strong case for moral symmetry.

Arguments for moral differences have been based on factors other than probabilities, however, and we must examine these.

III

We may now deal with an issue I noted briefly at the end of part I—whether there is not, underlying the pairs on my lists, an important asymmetry. Is the kind of “bringing about” in refraining from movement really comparable to that of moving? If a movement “brings about” a probability of 1 of some ESS, the assumption is that in the absence of that movement the probability would not have been 1, indeed that the ESS would not have occurred. By contrast, if a refraining is held responsible for the probability of some ESS being 1, then the assumption must be that the refraining occurred in a situation in which other factors already present would make the probability 1, *unless* some movement is performed from which the agent then refrains. In order to say that the probability would not have been 1 in the absence of that refraining, the “absence” must be interpreted as the actual making of the movement from which the agent refrained.

All of this is true enough, but it is difficult to state it as an argument without all the questions being begged. If one is already questioning the moral significance of the distinction between moving and refraining *as such*, it will not be very persuasive to point out that an analysis of the situation in which a movement brings about an ESS and of the situation in which a refraining brings about an ESS reveals that in the former case the absence of a movement would have made the difference whereas in the latter the absence of a refraining would have made the difference. This is the same distinction being put to work which was under question.

One point can be made, however. It is not at all unusual for people to deliberately perform movements which have as an unexpected result some ESS. On the other hand, it would be rather sur-

prising if someone refrained from a movement without realizing the difference the movement would have made to some ESS. This is because the mere absence of a movement is not a case of refraining—one must have considered the movement and decided against it. And one would want to know why, of all possible movements, just that one was under consideration if one did not realize its significance. Still, such situations are possible. Suppose there is a lever on a post in a field. It so happens that if someone pulls the lever by a certain time it will disconnect a circuit and prevent an explosion which will kill hundreds. A man crosses the field and sees the lever. He is tempted to pull it just for the fun of it, but decides not to and goes on his way feeling rather virtuous, though in fact by so refraining he has failed to prevent a disaster. Perhaps we can say this much—that with refrainings one is more likely to fulfill the knowledge conditions needed for responsibility than in the case of movements. But it is the presence or absence of the relevant knowledge which then makes the moral difference, not whether it is a case of doing or refraining.

It is likely, however, that issues of responsibility, and of credit and blame, are at the heart of many of the distinctions people wish to make. We may see this informally by considering a pair from the positive half of my list, say 1 and 2, in which some valuable state of affairs results from a doing or from a refraining. Surely, one wants to say, the person who actively brings about some benefit deserves more credit for it than the person who merely refrains from interfering with a state of affairs which is already producing the benefit. This seems intuitively plausible, but a little thought should call it in question. To take a simple case. One person spends an hour putting up a bird house which helps bring about a successful nesting of bluebirds (who are not doing so well these days). Another discovers bluebirds nesting at the hinge of his closed garage door. Rather than disturb the birds, he leaves his car in the garage for the duration of the nesting period, at considerable personal inconvenience. Who, here, deserves more credit? There is no universal rule that doing requires greater effort than refraining. Amount of effort is the morally relevant variable, not whether it is attached to doing or refraining.

What of more direct issues of responsibility? For these, let us look at another of Trammell's arguments, in which he claims that “Still another factor underlying the distinction between negative

and positive duties is responsibility."²² He deals with cases involving some NESS—specifically, the loss of human lives through killing or failing to save. He asks us to consider who initially places a life at risk so that it requires saving.

In general, if x kills y , then x is responsible for y 's death. But if x fails to save y , then x may or may not be responsible for x being in the situation in which y needs to be saved. The more directly involved x is for y 's needing to be saved, the more responsible x is for helping to rescue y Finally, a person is not necessarily responsible for someone else's needing to be saved; but he is responsible for the life of anyone he kills.²³

There is certainly a moral point here which I would wish to acknowledge. I do want to say that the industry which is polluting a river should be held responsible for the harm that pollution causes and that the prevention of that harm should focus on its source, rather than requiring cities downstream to remove the junk that should never have gone into the water in the first place. Having said this, however, could one go on to argue that once the pollutants are in the river the cities are not responsible for removing them before passing the water on to their citizens? And if they are not acting, what should private citizens do?

This kind of case is particularly instructive and shows, I think, why Trammell's argument seems so morally unresponsive. We are trying to determine the responsibility of agents who find themselves in a situation in which they could save lives which are at risk for *whatever cause*. To say that they are not responsible for those lives being at risk seems, at that stage, beside the point. Furthermore, as pollution cases tend to illustrate most dramatically, those initially responsible for creating a risk are often least likely to do anything about it. They generally derive some economic or other benefit from the activities creating the risk and thus have a positive disincentive for doing anything. So while we may indeed say they ought to be doing something, we must also, realistically, say that they probably will not. (Since the Environmental Protection Agency issued its report on the large number of carcinogenic substances in the drinking water of New Orleans, it has been the Environmental Defense Fund and the Ecology Center, two private conservation organizations, which have

pressed for action, not the city officials and certainly not the petrochemical industries upriver on the Mississippi.)

When, then, we consider an agent faced with an existing risk situation, what he must take into account is not what others ought to do but what they are likely to do. At that point they form part of the probability matrix within which he must calculate the results of his decisions. If we wait for the initiators of risk (and who are they in natural disasters?) to act, the harm will be done. If we wait for the Japanese and Russian whalers to institute proper restraints, the whales will be gone. Looking at disastrous end results, if we know we can do something to prevent them can we in good conscience say we are not responsible?

The environmental cases I have mentioned are not really on my table, of course. There is no single doing or refraining which could reduce the probability of the impending harm to zero. The most one can ask of an agent is to do what he can to lower the probabilities, but that is also the least one should ask.

But does this not place on the average moral agent an impossible burden? How much can we morally *require* of people? This sort of argument for "moral realism" is a final line of defense which we must now examine.

IV

How strenuous should morality be? We may introduce this issue with a final argument by Trammell to the effect that "Denial of the distinction between negative and positive duties leads straight to an ethic so strenuous that it might give pause even to a philosophical John the Baptist." He states the argument as follows:

A number of factors underlie the distinction between negative and positive duties, one of which is the dischargeability of a duty. It is an empirical fact that in most cases it is possible for a person not to inflict serious physical injury on any other person. It is also an empirical fact that in no case is it possible for a person to aid everyone who needs help. The positive duty to love one's neighbor or help those in need sets a maximum ethic which would never let us rest except to gather strength to resume the battle. But it is a rare case when we must really exert ourselves to keep from killing a person.²⁴

²² *Ibid.*, p. 135.

²³ *Ibid.*, p. 136.

²⁴ *Ibid.*, p. 133.

The general point, according to Trammell, is that "As a rule, fulfilling the duty not to kill leaves us free to carry out a vast range of activities without restraint."²⁵ He notes his agreement with a similar point made by P. J. Fitzgerald to the effect that the duty to refrain from a certain action "merely closes off one avenue of activity" whereas the duty to perform a certain action "closes off all activities but one."²⁶

The first point to be made is that Trammell and Fitzgerald are here focusing on cost to the agent, which is a consequentialist consideration and does nothing to show a moral difference between negative and positive duties *as such*. As I noted above, this *cost* or *amount of effort* is the morally relevant variable, not whether it is attached to doing or refraining.

At best, Trammell and Fitzgerald may be making a "pragmatic" point—that a distinction between negative and positive duties is a useful rule because *in fact* positive duties are generally more strenuous than negative.

There is a critical assumption behind such arguments, namely that there is a vast range of our ordinary daily activities which are quite innocent both in intention and—more to the point—in result. This may have been roughly accurate in earlier times (though this is a complex historical question), but it is a dangerously misleading assumption in our world today.

Let us focus on a type of result which is certainly of great moral concern—the loss of lives. If one is thinking of the traditional one-on-one shooting or stabbing as the relevant type of action, then the assumption may seem justified. Such actions fit into my preliminary table, in which we restricted our probabilities to one or zero (essentially those involving a fairly direct causal line from the implemented decision of a single agent to the ESS).

In order to have a table which even begins to do justice to our world today, however, we must consider a much wider range of probabilities and we must place the ID of one agent in the context of the IDs of all other agents. What we must take into consideration is seldom the much debated "What *if* all (or many) others were to do the same" but instead "Since many others *are* doing the same . . .". When we place our actions in the actual

interconnections of our world, how innocent are they?

I could go into a number of case histories here, but I can make my general point with two names—Paul Ehrlich and Barry Commoner.²⁷ Between them they represent the concern over two of the terrors of our times—population and technology. When ever increasing numbers consume the products of a technology which dumps into the environment increasing quantities of synthesized pollutants with which natural systems cannot cope, we have the setting for disaster. We can already document much of the loss of life—human lives in famine, killer smogs, and other visible disasters, other lives as whole species are eliminated at an accelerating rate. Much is hidden in public health statistics traceable to air and water pollution, or the accumulation of pesticides and other organic compounds in all living bodies. Beyond all of this is the growing concern with the supreme question. At what stage may we blindly go beyond the point of no return and destroy the very life-support capacity of whole regions or of the earth itself? How much pollution can we inflict on the oceans, which supply the great majority of our atmospheric oxygen? What are we doing to the ozone layer which helps screen us from ultraviolet rays? The list of questions is as endless as the stream of goods which now supports human life at all levels of consumption, but particularly those at the affluent end of the scale.

Under these conditions, in the developed countries of high technology innocent actions are becoming an endangered species. The most ordinary activities of consumption—eating our meals, driving our automobiles, buying our polyester clothes, cooling our homes—are meshing with the comparable activities of millions of others, with resources draining away at one end and pollutants piling up at the other. It is important to remember that in terms of various estimates, the average citizen in a developed country (of which the United States is the supreme example) places from ten to twenty times the stress on the environment as does a citizen of one of the poorer nations. One of the ironies in this is that it is precisely those who, in traditional terms, can "afford" to have them who cause the greatest harm by having large families.

²⁵ *Ibid.*, p. 134.

²⁶ "Acting and Refraining," *Analysis*, vol. 27 (1967), p. 136.

²⁷ Cf., Paul Ehrlich and Anne Ehrlich, *Population, Resources, Environment* (San Francisco, 1970); Barry Commoner, *The Closing Circle* (New York, 1972).

What is the proper moral response to what we are learning about the impact of our activities on all forms of life? If it is too much to ask that the ordinary citizen display the activism and austerity of a Ralph Nader—that “maximum morality” which Trammell views with such suspicion—what would be the appropriate “minimum morality”? One would at this stage get an ironic smile from anyone familiar with environmental literature. The phrase “minimum morality” has become common coin and has figured prominently in the debate over proliferation of nuclear power-plants. A representative quotation is the following:

Since the social requirements for acceptability of nuclear power are dominant and cannot be met, it follows that no group of humans has the moral right to support the construction or operation of nuclear power plants. *Minimum* morality, as many have stated, requires that we do not compromise the chance of life for generations to come.²⁸

Note that this is, in traditional terms, a “negative responsibility.” We are being asked to refrain from those activities which pose a serious threat to life. And all that this requires is that we alter radically our whole consumption oriented way of life, closing off vast ranges of activity we have come to take for granted.

It would be pleasant, of course, if moral theory could show us that we can pretty much go on in the same old ways—as decent, well-intentioned, law-abiding citizens. But it is not some aberrant moral philosophy which is destroying that dream, it is our increasing knowledge of how our lives are interconnected with the world.

University of New Orleans

In terms of my analysis, it is our responsibility to determine as best we can how any decision we might implement, whether a doing or a refraining, would affect the probabilities of the loss of lives and of the destruction of life-support systems. This moral requirement is the individual analogue of the legal requirement now placed on federal agencies by the National Environmental Policy Act—that before any project is initiated there must be a full study of its impact on the environment, natural and social. (And it should be noted that one of the alternatives which must be considered is “do nothing.”) We are slowly learning that this is the only way to avoid doing irreparable harm.

None of this will provide easy answers to any question. The relevant knowledge needed for this kind of decision making cannot be gained by the average person just looking about and using his common sense.²⁹ We need all the expert help we can get. The knowledge which is becoming current in environmental circles must somehow inform the lives of all citizens. In a fine phrase of Harris, we must “make people aware of what they are doing.”³⁰

The decisions we implement are in fact part of a complex matrix which has been producing the statistics of disaster at an accelerating rate. Once we know this, we have the sufficient conditions for sharing in the responsibility for those disasters. We may find that we have been paying an indefensible “life-price” for many of the material benefits of our way of life. Minimum morality may become very strenuous indeed.

Received December 15, 1975

²⁸ Dr. John W. Gofman in *The Case for a Nuclear Moratorium* (Environmental Action Foundation, Washington D.C.).

²⁹ Cf., Jay W. Forrester, “The Counterintuitive Behavior of Social Systems,” reprinted in R. T. Roelofs, J. N. Crowley, and D. L. Hardesty (eds.), *Environment and Society* (Englewood Cliffs, 1974), pp. 275–303.

³⁰ “The Marxist Conception of Violence,” *op. cit.*, p. 212.

IV. A PARADIGM OF PHILOSOPHY: HOHFELD ON LEGAL RIGHTS

THOMAS D. PERRY

WHY did F. P. Ramsey call Russell's theory of descriptions a paradigm of philosophy? Apparently because it does not merely tell us that a sentence like "The King of France is wise" involves "some possibly complex multiple relation between kingship, France, and wisdom," but rather "explains exactly what that relation is."¹ One could add that it neatly solves a philosophical problem: that of avoiding "Meinongian objects" foisted on us by apparently referential uses of "a unicorn," "the round square," "the present King of France," etc. At least it powerfully illuminates such problems of meaning and reference even if it can also be misleading in ways that Strawson has pointed out. To these criteria—exactness of analysis and "solving or illuminating power"—some writers would add that a model piece of philosophizing will contribute valuably to "the theory of critical discussion."² Perhaps it will improve our understanding of the nature of some type of critical discourse: science, ethics, literary criticism, whatever. Or perhaps it will actually contribute to the branch of critical thought and talk with which it is concerned by clarifying or reconstructing specific concepts, thus enabling moralists, scientists or literary critics, etc., to think more effectively. An especially admirable work may even do both of these things.

I wish to show that by these same criteria W. N. Hohfeld's analysis of legal rights³ is another great model of philosophy. We usually think of him as a legal theorist rather than as a philosopher because his interests in conceptual analysis were narrowly focused on law and far removed from what used

to be called pure philosophy. Yet—to take things in a somewhat different order than I did in the preceding paragraph—(1) Hohfeld explains exactly how several conceptions commonly called rights are complexly related; (2) he shows us how to recognize and avoid specific conceptual errors common in legal discourse by attending to these relations, errors that can easily result in practical mistakes, wrong decisions; (3) when we apply his theory to specific problems we discover something important about the nature of legal discourse, and especially judicial discourse, that lawyers and laymen alike have either not been aware of or have not clearly understood; and (4) a chronic philosophical problem about law is thereby much illuminated if not finally resolved. I shall adhere to this order of discussion in the four sections which follow.

I

Any discussion of Hohfeld will prudently begin by stating his scheme of relations, familiar as they are to many philosophers. On the present occasion, such a rehearsal will not only fix them in our minds for later use, but will also begin to establish my first point. An exact account of something is a precise and accurate one, and it will be obvious from the summary that Hohfeld specifies a rather precise set of conceptual relations which he says are involved in legal talk about rights. The considerable accuracy of his account will also become apparent as we go on to use it in later sections.

Hohfeld noticed that the idea of someone's

¹ Quoted from F. P. Ramsey by G. E. Moore in "Russell's 'Theory of Descriptions'," *The Philosophy of Bertrand Russell* (Evanston and Chicago, 1944), ed. by P. A. Schilpp, at 179. The passage quoted is at 142 of Ramsey's *Foundations of Mathematics and other Logical Essays* (London and New York, 1931), while the famous reference to "that paradigm of philosophy, Russell's theory of descriptions" is in a note on p. 263.

² For the thesis that analytical philosophy, if not the philosophy of sages, is properly understood as the theory of critical discussion, see John Passmore's article "Philosophy" in *The Encyclopedia of Philosophy* (New York, 1967), ed. by Paul Edwards. Of course, many recent writers have spoken of philosophy as a second-order discipline whose subject matter is not "the world" but the various arts and sciences or primary disciplines.

³ *Fundamental Legal Conceptions as Applied in Judicial Reasoning, and other Legal Essays* (New Haven, 1919, 1923). The first two essays presenting Hohfeld's analytical scheme were reprinted with the same pagination by Yale University Press in 1964 under the title *Fundamental Legal Conceptions as Applied in Judicial Reasoning*.

having a right to something is really a cluster of related ideas if we may judge from common legal uses of the word "right." Contrary to the saying that for every right there is a correlative duty, he pointed out that some rights have such correlates while others have not. Those rights which are correlative to duties he called right-claims (or rights in the strict sense). Mr. *A*'s right-claim that Mr. *B* shall do or forbear from doing *X* is correlative to *B*'s duty to do or forbear; they are two sides of the same coin. But sometimes when we are said to have a right to do *X* what is meant is that we are legally at liberty to do *X*, i.e., that no-one has a right-claim that we must forbear from doing *X*; or, if we are speaking of the legal relations of just two people, that the other person lacks such a right-claim. Thus, as between *A* and *B*, *A* may be at liberty to do *X*—that is, *B* will have no legal right in the strict sense that *A* shall not do it—but this does not mean that *B* is under any duty toward *A*. Hohfeld called such a liberty a "privilege," which is sometimes misleading for reasons we need not go into. Let us call it a liberty-right, as is now usually done, or simply a liberty. Next, there seems to be no convenient word for *B*'s corresponding relation to *A* when *A* has a liberty-right to do *X* so far as the two of them are concerned. Hohfeld called it a "no-right" on *B*'s part. A "no-right not" would be more literally correct but it is too awkward to adopt. That is, in the situation just mentioned *B* has no right that *A* shall *not* do *X*. Hohfeld called attention to this by remarking that a liberty and its correlative no-right are always of "opposite tenor."

So far this gives us two pairs of correlative terms: right/duty and liberty/no-right. Now each member of each pair has its logical opposite. Thus, if I do not have a right to *X*, I have a no-right to it; and if I do not have a duty to do *X*, then I have a liberty-right or am at liberty not to do it. Similarly for the second pair of correlatives: if I am not at liberty to do *X*, then I have a duty not to do it (as opposites, liberties and duties are also thus of opposite tenor); and if it is not the case that I have a no-right to *X*, then I have a right to it. In addition to rights and liberties, Hohfeld went on to point out and label two other "advantageous" legal relations often called a "right," namely a "power" (a legal capacity to change someone else's rights or other legal relations) and an "immunity" (an

immunity against having one's legal relations so changed by another person). Their correlatives and opposites are also listed in the note.⁴

When people are first learning these distinctions between different kinds of "rights," the most elusive one regularly proves to be that between a right-claim or right, and a liberty-right or liberty. We fail to retain it, I think, because at the outset we are not fully convinced that there is a real distinction here. After all, if I really have a liberty-right to do *X* or not to do *Y*, then surely other people *are* under a duty to me in this regard. Surely it is their duty not to interfere with my doing *X*, and it is their duty not to attempt to compel me to do *Y*. If they are legally free to interfere with me, how can I be said to have the liberty in question? Hohfeld anticipated this objection and took pains to explain that one can very well have a liberty without having any right not to have that liberty interfered with in this or that manner, or any manner. Depending on the circumstances, a given liberty may not be worth very much to us if it is not protected by certain right-claims, but the two relations are logically independent. To repeat his shrimp salad example, if I have paid for the salad I am at liberty to eat it, and you, my luncheon companion, have (correlatively) a no-right that I not eat it. Furthermore, I have a right not to be interfered with by you as I eat it. But I might *give* you the liberty-right to eat it, withholding any right-claim on your part not to be interfered with in eating it. Then if I do interfere with you in this regard (say, by quickly eating the salad myself) I violate no right of yours, no duty that I owe you, for you can only have such rights, liberties, powers and immunities with regard to my salad as I may choose to give you.

Even though the distinction between right-claims and liberty-rights is logically sound, it might be thought to have little or no practical significance. This is far from true, and Hohfeld's scheme of analysis gains much of its philosophical power from the importance of this distinction and the fact that we overlook it so easily. My illustrations will focus on this.

II

I turn now to my second point, that Hohfeld's scheme can directly improve the quality of legal discourse by saving judges and other lawyers from

Correlatives:	right-claim (right)	liberty-right (liberty)	power	immunity
	duty	no-right of opp. tenor	liability	disability
Opposites:	right-claim (right)	liberty-right (liberty)	power	immunity
	no-right	duty of opp. tenor	disability	liability

conceptual errors they are otherwise prone to commit. To approach the topic through an oversimplified example, consider the following case which I shall call the malicious injury case. *A* bribes *B* to discharge *C* from his job for no other reason than that *A* hates *C* and wants him to suffer. *C* learns why he was discharged and sues *A* for damages. The court awards him \$10,000, saying that everyone has a right to earn his living through lawful employment, and that third parties such as *A* are under a correlative duty not to interfere. This seems a natural thing to say, and it is not far from what courts have sometimes said in real cases. Now consider a second case which I shall call the nepotism case. *D* persuades his brother *E* to discharge *F* (a complete stranger to *D*) so that *D*'s son can be hired into the job, the only job in *E*'s firm for which the son has suitable skills. *F* learns why he was discharged and sues *D* for damages. This case arises in the same state in which the malicious injury case was decided earlier by the highest court of the state. Let us suppose that there are no other relevant precedents and no relevant statutes; also that a trial judge sitting without a jury has determined the foregoing facts. The question we have to consider (though we will not have to give a final answer to it) is, How should the judge decide the nepotism case?

Since the malicious injury case is a binding precedent, it might be thought that the judge is obliged to decide in *F*'s favor and order *D* to pay him compensation. After all, it is now the law of the state that a person has a right to pursue a lawful calling without interference; *F* was doing just that, and *D* who was under a corresponding duty not to interfere, did interfere. But the two cases differ in some respects, and those who are familiar with law and judicial practice know that a court is not obliged to accept and apply—precisely as formulated—the rules or reasons given by an authoritative court in a partly similar case. Although it is a civil wrong or tort to interfere with someone's lawful employment for purely malicious motives (and that was the sort of claim *C* had made and the high court had recognized, giving broader reasons than were necessary) this does not entail that it is a tort to interfere non-maliciously to help one's son find employment. No doubt *D*'s counsel will point this out to the court and argue, further, that such behavior did not violate *any* right of *F*'s. The judge too will probably recognize that in spite of the decision in the malicious injury case he is still free to consider whether the law does give people

the right not to be interfered with as *F* was interfered with—or rather, since there are no statutes or close precedents, whether the law *should* give people such a right.

Why is the judge in the nepotism case free to consider this? Here we find that Hohfeld's distinctions are useful both in stating the law which the judge must take into account, and in understanding the kind of intellectual problem which he faces. First as to the existing law, i.e., the malicious injury case considered as a precedent. It is a binding precedent, as I have said, because it was decided by the highest court of the state (not, e.g. by a prestigious court of some other common-law state or country) and is in some ways similar to the nepotism case. But how binding are the liberal reasons or rule formulations found in it? They are not only broader than they needed to be, as I said before; they are logically incoherent, and seeing this can help us appreciate the actual force of the decision as law. We may safely assume that under the law as it existed before the malicious injury case it was every person's right (in the sense of a liberty) to pursue a lawful calling; *that is*, no one else had a right (in the sense of a right-claim) that any given person should not pursue a lawful calling (except perhaps by private agreement between them). This latter no-right is the true correlative of that undoubted liberty; we cannot infer as its correlative that anyone else is under a duty not to interfere. The court in the malicious injury case therefore needed some other reason than the fallacious inference just mentioned to support its statement that *A* was under such a legal duty. Since it gave no other reason, about the best we can do in assigning a reason for the decision is to assume that the court was of the opinion that malicious behavior like *A*'s is so morally objectionable or socially harmful if tolerated that the law ought to protect people from it by giving them a civil remedy against it.

Now if *that* is the reason why *C* had a right not to be interfered with, it does little to support *F*'s claim in the nepotism case. Is it morally objectionable to persuade one close family member to help another? Perhaps it is, in the circumstances described and with the effect mentioned. But it is obviously less objectionable than what *A* did, and one might plausibly think that this sort of behavior would have fewer evil consequences if tolerated by law than behavior like *A*'s would have. The judge in the nepotism case is thus faced with an essentially legislative issue; there is no law on the specific point he has to decide. And the reasons formulated

by the court in the malicious injury case not only fail to give him any guidance on how to resolve that issue, they are a snare that can keep him from seeing that he has such an issue—a lawmaking or “policy” issue—to decide, so that he may end by deciding it blindly. If he simply applies that formulation as the major premiss of a practical syllogism (the minor being that *F* was pursuing lawful employment with *E*, and *D* did interfere) he will not at all have considered why the law should protect us against economic interference of *this* kind and motivation, or why it should not. Few judges would be quite that mechanical about it; they would also probably feel that in the special circumstances of the nepotism case *F* ought to be compensated by *D* (or that he ought not). But mere feeling is no substitute for the reflective policy making which the case actually calls for. If a judge of the former conviction should decide for *F* on the basis of the formula produced in the earlier case, this would only serve to obscure the need for such reflection.

In the well-known case of *Quinn v. Leatham*,⁵ the judges of Britain’s highest court made a rather similar mistake. They failed to recognize a policy-making issue in the same area of tort law (economic competition) and they did so by confusing liberty-rights with right-claims. The defendant Quinn and other union organizers had attempted to force Leatham, a butcher, to replace his employees with union members. For this purpose they threatened to call a strike at the shop of one of his customers (another butcher) unless the latter would stop buying meat from him. The customer complied with this demand, causing financial loss to Leatham who brought this action and won a verdict for damages. When the House of Lords upheld this result, Lord Lindley gave the reason that Leatham, like anyone else, had a right to pursue whatever lawful business or employment he chose, and that other people, including the defendants, were under a correlative duty not to interfere with his business without lawful justification.

In commenting on the case, Hohfeld⁶ complained that Lord Lindley’s reasoning confuses the plaintiff’s undoubted liberty to carry on a lawful business with a right-claim not to be interfered with by others when doing so. That liberty entails (has as its correlative) only that the defendants had no right that plaintiff *not* pursue a lawful business, while such a right-claim on his part would entail

that they had a duty not to interfere. Whether there is or should be such a right and correlative duty protecting the plaintiff’s liberty is a separate question of legal policy to be considered on its merits and not to be thus begged in a specious inference from a principle of law no one would or could question, i.e., the principle that everyone is at liberty to pursue a lawful calling. Thus it would seem that Lord Lindley and the unanimous House managed to overlook the real issue of the case—blindly deciding it nevertheless—by confusing these two senses of the word “right.”

Hohfeld’s criticism oversimplifies things a little because Lord Lindley did not purport to infer an absolute duty of noninterference but a duty not to interfere without lawful justification. He and the other Lords held that there was no such justification in this case. They were impressed by the fact that Leatham had offered to run a union shop if he could pay his men’s initiation fees and “fines” in the union and keep them on, but that the defendants had insisted that the men be replaced by others who had joined the union earlier. They said Leatham’s men would “have to walk the streets for a year.” This seemed pure malice to some of the judges; Lords Macnaghten and Shand said repeatedly that the defendants acted solely to injure the plaintiff and his employees rather than to further their own interests.

It was an error, nevertheless, for Lord Lindley to infer *any* duty as the strict correlative of the plaintiff’s right to engage in business, for that right was only a liberty. He and some of the other judges mistakenly thought that one man’s liberty to engage in a lawful calling is equivalent to a duty on everyone else’s part not to interfere except by lawful means. The two things may sound similar but Hohfeld’s scheme makes it clear that they are logically independent. Now it did happen to be the law in Britain that a person’s liberty to engage in a lawful calling *was* protected by right-claims not to be interfered with in various ways by behavior which was unlawful under specific principles of tort law, e.g., by the intimidation of others to harm him economically, or by a “conspiracy,” or by inducing others to break their contracts with him. But whether it was also the law that one who interferes with the economic affairs of another to the latter’s loss, by conduct not unlawful under any of those principles, can be made to pay compensation unless he shows that he acted with just

⁵ [1901] *Appeal Cases*, 495.

⁶ *Op. cit.*, 42–43.

cause or excuse, was quite another question. And it certainly was not to be answered on the basis of a fallacious inference of a duty from a liberty.⁷ The House is also open to the criticism that even if the defendants did have such a duty not to interfere without lawful justification, little attention was given to considerations which might tend to justify or excuse their actions. More on this later.

So much then for the second point I wished to make in arguing that Hohfeld's theory of rights is a paradigm of philosophy: that it can definitely help judges and other lawyers avoid conceptual errors which they have often committed. It can do so because it explicitly marks distinctions and relations which must be recognized if such errors are to be avoided. Thus it certainly improves the tools which people have to work with who engage in this particular type of critical discourse.

III

My third point was that when we apply Hohfeld's scheme to specific problems we also gain a better understanding of the nature of that type of discourse. What lesson can we learn from the foregoing illustrations about the nature of judicial decision making and justificatory discourse? For one thing, we have just seen that decision making is sometimes law making; that it is not always or wholly a matter of finding and applying existing law and explaining how this was done. Sometimes a judge can hardly *avoid* making law, as illustrated by the nepotism case. But I suppose we have all learned this at our mothers' knees, and although Hohfeld deserves part of the credit for enlightening our mothers on the point, I prefer to emphasize certain related but deeper points which are not well understood even now. What sort of questions does a judge need to ask himself when deciding *whether* to legislate in cases where he has a choice? How is he to perform this part of his function intelligently? And why is good judicial lawmaking

never simply "result oriented"? That is, why is it not enough for a judge to consider what outcome would be most just or sensible in the case before him, and then reshape the existing law to accommodate it? I shall attempt to explain very briefly how Hohfeld gives us valuable insight on these matters.

When we apply his scheme to *Quinn v. Leatham* we see how legal analysis may be required before a judge can even understand *what* the legislative or policy issue is in the case before him. That issue is never to be understood merely by asking broadly whether the law ought to give relief to people situated as the plaintiff was situated, against people who have behaved as the defendant behaved. And it is only when we see the more specific issue that we can give proper attention to the considerations which are relevant for resolving it one way or another. What then was the legislative issue in *Quinn*? There really were two such issues, one a rather broad and doubtful question of social policy, and the other a much narrower question which arises when the first has been answered affirmatively.

- (I) Should everyone have the duty not to interfere with another person's lawful economic activities and arrangements, to that person's economic loss, except with "just cause or excuse"? and if so,
- (II) Should the court hold, and so establish as a precedent, that interference of the sort engaged in by *Quinn* and the others was without just cause or excuse?

We have seen that the court answered (II) in the affirmative, as though (I) had already been answered Yes, whereas in fact (I) had not even been considered because of the specious inference that the duty to which it refers was already legally established. Now even to begin to consider (I) intelligently, the court would have had to pay attention to the fact that the liberty to engage in lawful economic activity was already protected by

⁷ As H. L. A. Hart comments on *Quinn v. Leatham*, *Allen v. Flood*, [1898] A.C. 1, and *Mogul Steamship Co. v. McGregor* [1892] A.C. 25, the House was faced in these cases with the question "whether an individual's liberty-rights to trade or employ labour or sell his labour are protected by a perimeter consisting only of duties corresponding to the specific torts of conspiracy, intimidation, and the inducement of breach of contract, or by a perimeter consisting also of a duty corresponding to the more general tort of interfering with the trade, business, or employment of a person without lawful justification or excuse." *Oxford Essays in Jurisprudence*, 2nd series (Oxford, 1973), ed. by A. W. B. Simpson, 181, n. The courts of several American states have adopted this more general, *prima facie* tort, as we may call it, often citing one or more of the above cases as persuasive authority. See W. L. Prosser, *Handbook of the Law of Torts*, 3rd ed. (St. Paul, 1964), 26, and the cases cited there. While a court could have substantial reasons for creating or recognizing this *prima facie* tort, as discussed later in my text, many of the judges involved in these leading British cases anticipate Lord Lindley in relying on the fallacious liberty/duty correlation. Several quote with approval Sir William Erle's treatise *The Law Relating to Trade Unions* (London, 1869), 12, where this highly seductive fallacy seems to have originated.

various specific duties of noninterference, as mentioned above. It would have had to consider whether there was a need to add this further protective right and duty if people are to be free of undesirable interference. And it would have had to consider whether this added protection is worth the price of imposing so vague a duty on economic competitors. Is it desirable or just that people should be made to pay court judgments when they have competed in ways they thought were within the rules, considering that they are given no better guide as to what the rules are than "just cause or excuse?" And would imposing this new duty place too severe a damper on competition, assuming that vigorous competition is socially desirable? Or would it perhaps tend to increase each person's freedom from interferences while placing a desirable brake on competition which we may believe has many ill effects if left unchecked? These are questions of social fact and of the relative desirability of different conditions of life, questions which legislatures may be in a better position than courts to answer, or which they may have better political standing to answer. If the law Lords had seen that they needed to address themselves to issue (I) as a law-making issue and so had come to consider these latter questions, they might well have concluded that the problem was one for Parliament and not for them. Thus they might have decided the case in *Quinn's* favor on the ground that there was as yet no such duty, no such *prima facie* tort as that referred to in (I), and that *Leathem* had failed to prove that *Quinn* and the other defendants had committed any of the previously established torts. But if the Lords had consciously accepted the challenge to legislate on (I), they would perhaps have seen that they needed different and more specific reasons for disposing of it than the following reason given by Lord Chancellor Halsbury:

"If upon these facts so found the plaintiff could have no remedy against those who had thus injured him, it could hardly be said that our jurisprudence was that of a civilized community . . ."⁸

This statement, so typical of the feeling expressed by all the Lords, is more immediately relevant to issue (II) than it is to (I). Along with the mistaken inference which prejudges (I) and focuses our whole attention on (II), it is really the principal reason the Lords gave for their decision. This is judicial lawmaking of the least self-conscious and

circumspect kind. It misses one major policy issue entirely (deciding it nonetheless) and although it does consciously decide the second policy issue, it does so without realizing that this issue depends on the first. Furthermore, its consideration of the second issue is very blunt and uncritical in the way it reaches out for a desired result without giving any specific reasons. It says no more than that the defendants' behavior was civilly intolerable in the Lords' opinion.

To sum up this third part of my argument, when we recognize these deficiencies in the Lords' performance we not only realize that judges make law, but we understand more of the problematics of their lawmaking. We get a better idea of the sort of analysis that has to be made if policy-making issues are to be clearly identified, and the sort of questions that have to be considered if such issues are to be decided intelligently. It is Hohfeld's scheme which has enabled us to pinpoint these deficiencies and gain these insights.

IV

"At least the judges in *Quinn v. Leathem* arrived at the right result (upholding the verdict for *Leathem*) even if their reasoning was poor." I believe many lawyers and laymen alike would agree with this statement, which I shall make use of in discussing my last point. This was the claim that Hohfeld's theory of rights is also a paradigm of philosophy in that it helps us solve one or more philosophical problems, or at least throws important light on them. I am thinking especially of the problem whether law is fundamentally distinct from morality, as "legal positivists" maintain while recognizing a genetic relation between the two, or whether law and morality are at bottom indistinguishable, as "natural law theorists" insist while admitting that there are superficial differences between legal and moral institutions. I shall briefly indicate just one of the many forms this problem has taken in the discussions of philosophers.

The variation I have in mind focuses on Anglo-American law and arises from a conflict or apparent conflict between two well developed and separately plausible positions. On the one hand, there are many reasons to think that law and morals are distinct in societies like ours which possess a fully developed legal system. In such a system the traditional rules of behavior which order the life of a

⁸ [1901] A.C. 506.

primitive society have been supplemented and partly replaced by (a) rules for changing and creating rules of behavior, (b) new rules so created, (c) rules for the adjudication of disputes, (d) rules for the enforcement of other rules, and (e) a more or less complicated rule for determining which rules are legally obligatory or officially cognizable, and which are not.⁹ This historical development drives a conceptual wedge between law and morals in ways that can only be briefly mentioned here. For one thing, rules of the legal system can be created, modified, abolished, while it seems absurd to speak of deliberately introducing or changing the rules of morality. For another thing, some recognized rules of morality never do have close counterparts in the legal system, and the similarity between some other moral and legal rules is more verbal than real. Finally, there is no logical reason why some rules of the legal system cannot be contrary to morality, whether we are thinking of popular or critical morality. Indeed, it is notorious that this has often occurred.

On the other hand, it is well known that Anglo-American judges often fall back on moral considerations when introducing and "developing" legal rules and principles; they are even expected to do so. It has often been said that, in deciding difficult or novel cases, our judges cannot really separate the law as it is from the law as it ought to be. They cannot distinguish between what the law *does* provide on the point at issue and what it *should* provide. And this does not simply mean that the courts sometimes have to make law on questions which the existing statutes and precedents leave unsettled. It also means—so the argument goes—that the law as it is cannot be separated from the law as it *really* ought to be. Often there will be only one right way to make law, one right way to decide the novel case at hand and lay down a precedent. It is always the judge's duty to decide "according to law," as we say, but this means according to binding statutes and precedents when there are such authorities controlling the case, and *otherwise* according to sound morality. A judge is not free to decide a case any way he pleases in the absence of specific binding legal standards. To elaborate this a bit, when we say that a judge is obliged to render the *best* decision in such a case, we do not mean the best

according to absolute morality, if there is such a thing, but neither do we mean the best according to any moral standards he may happen to embrace, however eccentric these may be. We mean the best according to common or social morality, the morality of his society. It is this morality which he has the duty to implement. We may call this a legal duty since it attaches to his role in a legal institution, but it is itself a duty of social and political morality. It is a judge's moral duty to follow the established practices of the courts in this respect as in others.¹⁰

This is why we can say—to complete the argument—that the House of Lords did reach the right result in *Quinn v. Leathem*, whatever the quality of the reasoning by which it sought to justify that result. For, in the first place, the relevant precedents and statutes did not clearly determine the result, as anyone must agree who studies all the speeches in this case and in *Allen v. Flood* and *Mogul Steamship Co. v. Macgregor*. And in the second place, common moral opinion would certainly hold that the defendants committed a great wrong against the plaintiff and his employees, who ought to have legal redress. This outcome was the product of a moral judgment which the court was legally and morally required to make. Thus it illustrates the point that legal and moral standards are indistinguishable in Anglo-American law, at least in doubtful cases.

Now how can Hohfeld's theory of rights help us resolve or clarify this problem? For a start, let us provisionally agree that the court did reach the right result in the *Quinn* case. This is an initially plausible example of that "sound morality" referred to in the foregoing argument: what people in our society normally think is morally right, or would think is morally right if they stopped to consider the matter. Or more to the point, let us provisionally agree that this is a sample of the common morality of Britain at the turn of the century when this case was decided. But notice that if we say that judgment for the plaintiff *was* the right result for the court to reach, this is tantamount to saying that the answer to issue (I) and issue (II) of the preceding section is Yes in both cases, as previously noted. It is important now to explain why this is so.

No one can plausibly hold that judgment for the

⁹ See H. L. A. Hart, *The Concept of Law* (Oxford, 1961), esp. Chapters 5 and 6.

¹⁰ Cf. R. M. Dworkin, "The Model of Rules," *University of Chicago Law Review*, vol. 35 (1967), pp. 14-46, reprinted in *Essays in Legal Philosophy* (Berkeley, 1968), ed. by Robert S. Summers; and R. M. Dworkin, "Social Rules and Legal Theory," *Yale Law Journal*, vol. 81 (1972), pp. 855-890. There are many arguments in these two essays which can be used to support the position summarized in this paragraph of my text. But see note 15 below.

plaintiff was the right outcome because of some consideration which applies, and can only apply, to this one case. Legal and moral decisions are supposed to be principled, and the only question is how wide or narrow the principle should be. Now it is conceivable that a court could devise some new principle of tort liability under which behavior like Quinn's and his associates would be specifically wrongful, thus adding to the "protective perimeter" surrounding the liberty to engage in a lawful calling, but doing so without laying down the broad principle of prima facie liability referred to in issue (I). But while this is logically possible, it is impractical as legal technique. If courts thus began to add piecemeal to the specific torts protecting this economic liberty, they would seem to be engaged in a rather futile undertaking. For in arriving at each new tort definition they would have to ask themselves whether certain behavior interfering with the economic activities and arrangements of others is justifiable or excusable—the very same test that would be employed by the broad principle stated in issue (I). Furthermore, conceptions of fair competition seem to change as socio-economic conditions change, so that these specialized principles of tort liability might have to be revised fairly often. Better then to establish the broad prima facie tort if we are going to add to the protective perimeter at all. Indeed, the appellate judges of Britain had already begun to speak as if there were such a prima facie tort,¹¹ although no one before Quinn seems to have been found guilty of it and its existence had been based on the mistaken inference we have repeatedly criticized. These are the reasons why, in saying that the court reached the right outcome in *Quinn*, we are saying in effect that the answer to issue (I) is Yes, and therefore that the answer to issue (II) is also Yes because the rightness of the outcome *then* requires that the sort of behavior these defendants engaged in *was* without just cause or excuse.

Now as soon as we are clear that these two issues are involved, so that we direct our attention to both of them rather than "presupposing" an answer to one and failing to distinguish between them, it be-

comes *doubtful* how they should be answered. The broader and more fundamental issue, (I), is especially doubtful for reasons already pointed out: the fact that such a legal restraint and ground of liability is extremely vague and therefore possibly unfair; the need to inquire into its probable economic consequences; the fact that people might easily disagree as to the desirability of those consequences. Even issue (II) will at least become debatable if we actually stop to think about it, perhaps freeing ourselves for a moment from the kind of class bias that dominated the Lords' thinking. Certainly it is questionable that the defendants acted solely out of malice, as some of the Lords said. By taking such a hard line with Leatham and his employees, Quinn and the others may have hoped to bring pressure throughout the industry, pressure on workers to join the union promptly and on employers to encourage this and to deal with the union. Moreover, the union may well have had the long-run objective of improving the lot of butcher workers generally, not merely that of its present members. While some of the Lords did mention this possibility, they did not seriously consider whether the means adopted might be justified by such economic goals. This is especially remarkable in light of the then recent *Mogul Steamship* decision, where the House had allowed as non-tortious some rather fierce competitive tactics between businessmen as long as their behavior did not fall under one of the older specific torts. Reading *Quinn* and *Mogul Steamship* together, one cannot escape the conclusion that it simply never occurred to the Lords that standards of fair competition between entrepreneurs might also apply as between capital and labor, or as between competing groups of workers. Not that the identical standards necessarily do apply to these different categories of competition, but the Lords can hardly be said to have had *any* concept of fair competition between labor and capital. Otherwise, it is difficult to understand how they could have failed even to consider whether harsh tactics might be as tolerable there as they were held to be in competition between business firms.¹² At all events, it is by no means

¹¹ See note 7, above.

¹² In the *Mogul* case the defendant steamship companies had combined to establish below-cost rates and preferential rebates to shippers in order to drive other carriers out of certain China ports and enjoy an effective monopoly thereafter. While this was intended to injure, and did injure the plaintiff companies economically, the Court of Appeal and the House of Lords held that it was justified by the defendants' pursuit of their own economic interests without personal ill-will toward their competitors, or any desire to injure them beyond the extent necessary to further their own economic goals. Obviously Quinn *et al.* could have argued to similar effect: that their union did act to injure Leatham and his employees economically as a means of obtaining economic advantages for themselves, but without personal ill-will, etc., and that they acted not only to further their own economic interests but ultimately the interests of all persons employed as butchers. The Lords tended either to overlook this

obvious that issue (II) should have been answered in the affirmative, and issue (I) is even more doubtful.

Does all this refute the claim that when there is a morally best outcome in a case not controlled by statutes or close precedents, that outcome is "required by law"? Obviously not, but I think it can give us a more adequate view of that claim and its alleged consequences for the relationship between law and morality. The discussion tends to undermine any assumption that there usually is a morally best outcome to legally doubtful cases, an outcome whose moral superiority is plain enough that it becomes the judges' duty to reach it. It is this assumption, I submit, which gives color to the formula that, in doubtful cases, what the law provides cannot be distinguished from what it really ought to provide, and to the resulting theoretical claim that law is indistinguishable from morality. The tacit suggestion is that the courts can readily identify what the law really ought to provide on some newly contested issue, and then fit this into the pre-existing system of law. But that can seldom be done, even if we concede that a sufficient index of what the law "really ought to provide" is moral consensus in the community. When new legal issues arise from new economic activities and associations—new forms of life, in Wittgenstein's phrase—there is not likely to be a consensus. We are more likely to find moral confusion and disagreement, if not the "battleground" Mr. Justice Holmes used to speak of in this connection. The law of labor-management relations, so painfully developed in the 75-year period since the *Quinn* decision, is itself the most vivid illustration of this. Even if judges do not take their own moral views as the criterion of what the law ought to provide, and do not appeal to a fictitious consensus, but actually attempt to estimate the relative advantages of having the law this way or that, the issue will still usually be controversial.¹³ For even when it is safely determined (as it seldom is) that such and such advantages to such and such people and interests will accrue

from having the law provide so and so, and such and such other advantages to such and such other people and interests from having the law provide something else, the question remains: Which of these people and interests ought to be favored? There is not likely to be a consensus on this question either, and no ethical theory is canonically for answering it. Nor are we to suppose naively that the interests of all will be advanced equally by some new rule of law, or that it would be generally accepted as just that they should be, given existing inequalities.

Quite typically, then, there is no "best" decision which the judges have a duty to reach in the doubtful cases to which the formula applies. The judges simply have to make up their minds as best they can, no doubt within the *limits* of existing social morality where these are discernible and relevant. The formula should be taken to mean not that law and morals are indistinguishable but merely that decision making does sometimes involve law making.¹⁴ The law made within those limits is the law whatever its wisdom or moral quality; but more than that, we have to recognize that the judges are not legally bound beforehand to reach the morally best result, because there is no morally best result. That is the situation in which judicial lawmaking normally occurs, even if we cannot say that cases never arise in which only one way of resolving an otherwise doubtful legal issue is consistent with contemporary morality. It can happen that there is a consensus on such an issue, or that no representative moral opinion found in the community would accept any other way of resolving the issue if the question were popularly explored. Perhaps in these contingencies it would be a judge's institutional and "legal" duty to reach this "morally best" outcome, at least in Anglo-American practice. But it is naive not to realize how unlikely these contingencies are. When we are theorizing about the nature of law, it is therefore a mistake to attach very much weight to what judges are supposed to

possibility entirely or miss its considerable weight in light of the *Mogul* case. Thus, Lord Lindley: "That they acted as they did in furtherance of what they considered the interests of union men may probably be fairly assumed in their favor, although they did not come forward and say so themselves [the defendants called no witnesses]; but that is all [*sic*] that can be said for them." (1901 *Appeal Cases*, at 536.)

¹³ Cf. Mr. Justice Holmes in "The Path of the Law": "I cannot but believe that if the training of lawyers led them habitually to consider more definitely and explicitly the social advantage on which the rule they lay down must be justified, they sometimes would hesitate where now they are confident, and see that really they were taking sides upon debatable and often burning questions." *Harvard Law Review*, vol. 10 (1897), 457, at 468. Reprinted many places.

¹⁴ The formula also reflects the fact that judges and other lawyers tend to speak of new judge-made law as though it existed all along. ("We hold that the plaintiff had a right to so and so.") For an illustration, notice how Hart, who is under no illusions in such matters, uses the present indicative in the passage quoted in note 7 above. If we take him too literally, we will understand that the question confronting the House was not whether the protective perimeter *should* and thenceforth *would* include the general *prima facie* tort (with retroactive effect for the litigants), but whether it already did.

do on these rare occasions. We are dealing here with an exception to an exception after all. In discussing the nature of law in relation to morality, we ought to remember that the vast majority of legal questions arising in daily life have definite answers in the existing statutes and precedents, as do most of the legal questions which figure in litigated cases, although the disputed legal *issues* may not. And few of the legal questions which are not thus definitely answerable have "morally best" answers in the relevant sense of the phrase which we have just been discussing.¹⁵

To conclude our discussion of this fourth point, let us glance back to see how Hohfeld's theory of legal rights has enabled us to gain this insight into our philosophical problem. The story began in earlier sections and culminates here. It was only because we were able to use his scheme to identify the policy-making issues in our principal case that we were able to focus our attention on them and ask whether there is any uniquely correct response to them as a matter of social morality. When we saw that there is not—and why there is not—it became obvious that this will be typical of most policy-making issues coming before the courts. When courts seek to resolve such issues, the consentient standards of social morality can seldom function as binding legal norms for the simple reason that they are logically unavailable, they fail to provide an answer. To be sure, when a judge tries to make law as wisely as he can, he may well invoke a moral standard, i.e., his own standard or that of some section of the community. But he is not legally obliged to invoke and follow *it*, i.e., that particular standard, even if it is his legal duty sincerely to apply *some* moral standard or other. *It* is not the law on anyone's theory.

There is no need to restate our four criteria of a model piece of philosophizing and repeat summarily how Hohfeld's work satisfies them. I shall conclude instead with two remarks on what we can learn from his success about the role of philosophy in our larger intellectual life.

First, the idea still dies hard that analytical philosophy is too aloof from life and practice, and

that analytical jurisprudence in particular is a pedantic distraction from issues of social policy that normative and sociological jurists rightly emphasize. But if we do not merely "learn" Hohfeld's analytical scheme but try to apply it to real cases we may discover again how blind and haphazard normative thinking can be when it is not based on careful analysis. A judge may be as policy-minded and "activist" as anyone could wish, but if he misses the sorts of distinctions that Hohfeld has taught us to recognize he may never see *what* policy issues he has to make up his mind about, much less decide them intelligently. Meanwhile those sociological jurists or political scientists who think that legal analysis is a myth and that the result a court reaches is all that really matters, continue to misrepresent the judicial process at its historical best and fail to see how its average quality might be improved. These are points on which Hohfeld could be helpful to them.

Finally, a word about the perennial belief of philosophers that it is their business to give us insight into the hidden and the unknown—into Reality behind Appearance, into logical form beneath grammatical form, into unconscious errors and conceptual distortions yielding problems we have not known how to cope with. Everyone knows what mistaken and perverse theories have sometimes been generated in this tradition, but its continuing importance can scarcely be doubted. Hohfeld shows what valuable results of this kind can still be reached by a candid mind that has no taste for myth-making. In dealing with one sphere of critical discussion, he gives a convincing account of distinctions and connections commonly missed, how they are missed, and how we must proceed if we are to think more adequately. And he puts us in position to make sharp guesses about why some bit of blindness has afflicted those whose thought we are examining—and can still afflict us, as in the too easy assumption that the House of Lords at least came to the right conclusion. Thus his work illustrates a great theme running through the whole history of Western philosophy from the Presocratics to Russell and Wittgenstein.

SUNY/Buffalo

Received September 9, 1975

¹⁵ In the articles cited in note 10, above, R. M. Dworkin points out that courts sometimes refuse to follow applicable statutes or normally binding precedents. He believes that it is sometimes their institutional and moral duty so to refuse and employ moral principles as the correct norms of decision. The conservative moral-political principles of legislative supremacy and *stare decisis* carry great weight, he thinks, but they can be outweighed at times by other moral principles which also are binding on judges as judges. Thus Dworkin holds that legal and moral norms are *generally* indistinguishable, not just when a case is undetermined by valid statutes or close binding precedents. It is beyond the scope of my paper to discuss this variation of the problem whether law is distinct from morality.

V. KNOWLEDGE: THE CESSATION OF BELIEF

MERRILL RING

IN the *Meno* Socrates remarks "But it is not, I am sure, a mere guess to say that right opinion and knowledge are different." (98b) Chisholm comments on that remark "The distinction would seem to be obvious. If one has knowledge then one also has right or true opinion. But the converse is not true."¹ That comment has a parody in "Men and women are obviously different. If one is a man then one is also a woman, though the converse is not true." Chisholm's move must be attractive for it has a history. "Socrates: It now appears that they [knowledge and true belief] must be different things. Theatetus: Yes, Socrates, I have heard someone make the distinction. . . . He said that true belief with the addition of an account was knowledge." (201c-d)² In the absence of any general principle which leads from the distinctness of *X*'s and *Y*'s to the claim that to be an *X* is to be a *Y*, the reader wishes to know the warrant for connecting up knowledge and (true) belief in this way.

My interest in that warrant arises from the aim and methods of this paper. The goal is to provide a perspicuous representation of the relation between knowing something and believing it. There is extensive agreement that knowledge requires belief. I shall argue that those who hold such a view are wrong. The nature of, and correctives for, that misunderstanding will develop out of paying close attention to the warrants for the thesis that knowing is necessarily believing. Such is the justification for the zeal with which I will hunt down the mechanism of arguments supporting the thesis. A final program note: even the present main dissenter from that view, Colin Radford, attempts only to show, by means of counter-examples, that while there is strictly no entailment "in most [cases] belief is necessary for knowledge."³ That concession as well as his mode of argument reveal that Radford's is a minor heresy where the need is

to be pagan. I shall argue that knowing and believing are exclusive.

While several arguments have been produced in support of the idea that belief is necessary for knowledge, historically and logically only two are significant. The first of these starts from an ordinary situation in which both belief and knowledge are involved. That situation is most explicitly joined with its claimed consequences by A. D. Woozley.

Normally, the passage of thought is from belief to knowledge: one starts with belief and later (but not, of course, always) arrives at knowledge. That is, belief, when it is confirmed, becomes knowledge. . . . Knowing *p*, then, will consist of surely believing *p*. . . . Knowledge has thus been analyzed . . . as the limiting case of belief.⁴

Cases of passing from belief to knowledge have seemed to show, without further ado, that knowledge is a kind of belief. They show no such thing. In fact, Woozley's connective "That is" is, as things stand, a logical howler. "Passage cases" abound in which the original characterization no longer obtains when a later characterization is applicable. In the army one normally progresses from corporal to sergeant (or in crafts from journeyman to master) but a sergeant is not a kind of corporal, for he is no longer a corporal at all. Of course, these examples do not show that knowledge is not the limiting case of belief. For in other types of passage the original characterization is still applicable after completion of the process; e.g. one remains a B.A. upon receiving the further degree of Ph.D., a father remains a father upon becoming a grandfather. The point of the examples is that the fact of temporal transition from belief to knowledge does not establish the logical thesis that knowing is believing.

Counter-examples can be located too easily. That suggests there is an unannounced assumption which makes it seem obvious that passing from

¹ R. M. Chisholm, *Theory of Knowledge* (Englewood Cliffs, 1966), p. 5.

² Quotations of the *Theatetus* are from F. M. Cornford (London, 1935) and of the *Meno* from W. K. C. Guthrie (Baltimore, 1956).

³ Colin Radford, "Analyzing 'Know(s) That'," *Philosophical Quarterly*, vol. 20 (1970), pp. 222-229, see p. 229. Cf. also Radford's "Knowledge—By Examples," *Analysis*, vol. 27 (1966), pp. 1-22.

⁴ A. D. Woozley, *Theory of Knowledge* (London, 1949), pp. 179, 186.

belief to knowledge is like going from father to grandfather. To discover what completes that analogy let us look at other usages of the fact of passage in arguing the entailment thesis. Here Plato and Ayer are illuminating.

Meno asks Socrates (97d) why knowledge is more valuable than true opinion and what the difference between them is. Socrates answers both questions by an "analogy" (98b) with the "statues of Daedelus."

If you have one of his works untethered, it is not worth much; it gives you the slip like a runaway slave. But a tethered specimen is very valuable for they are magnificent creations. And that, I may say, has a bearing on the matter of true opinions. . . . They are not worth much until you tether them. . . . Once they are tied down, they become knowledge. . . . What distinguishes one from the other is the tether. (97c-98a)

Meno asked about the difference between knowledge and (true) belief. He receives an answer which emphasizes the tether, the differentia, and which thereby *presumes* that knowing is a species of believing. As Meno swallows that, it looks as if the anomaly of the *Theatetus* is here reproduced.

To see that it only looks that way, consider what Socrates is doing. Earlier the slave boy was brought from ignorance to true opinion and now Socrates is picturing the logic of the further temporal transition ("until you tether them," "once they are tied down") from true belief to knowledge. If Meno is here rationally convinced that knowing is a species of believing, he must be so because he accepts the fact of passage to show that. Why should he accept that? Is there anything to be found here which could serve to convince Meno that the passage from belief to knowledge is different from journeyman to master? If we are looking for an argument, we are bound to be disappointed. There is not a shred of one. But there is something on which Meno's acquiescence rests. Socrates' ideas that knowing is a species of believing and that passing from one to the other shows that to be so are contained in the story of the statues. Accept it, as Meno does, and you must accept everything else.

What that story amounts to is a representation of belief and of knowledge as analogically "like" objects, here statues. Why is such a representation appropriate? Why are we, like Socrates and Meno, so readily satisfied by the picture of belief and knowledge?

Plato is not alone in offering a picture in connection with passage cases. Though Ayer only suggests a story, what is suggested is fundamentally the same as Plato's representation.

One may also be completely sure of what one believes in cases where the belief is refused the title of knowledge on other grounds. . . . It is not hard to find examples of true and fully confident beliefs which in some ways fail to meet the standards for knowledge.⁵

There is a pictorialization of belief behind those words; for it is people, not beliefs, who are refused or granted the title of knowledge and it is people, not beliefs, who meet or fail to meet evidential standards for knowing. In writing as if those fortunes and calamities occur to beliefs rather than to people, Ayer suggests that passing from belief to knowledge is like what goes on at a dog show (a statutory contest would be equally appropriate.) There, various dogs (beliefs) are presented and judged with those meeting the standards being awarded the title "champion" (knowledge). Of course, even after becoming a champion, a dog remains a dog. Similarly, a belief remains a belief, though of an honored sort. Once again there is no *argument* that believing continues when someone passes on to knowing. Once again, a picture is held up in which belief and knowledge are represented as objects. The account of the concepts follows from, because it is embedded in, that.

The most expansive usages of passage cases give us (only) this much: to convince us that one still believes even when one has passed on to knowing, we are told stories in which belief and knowledge are represented as objects. The relevant questions are: why have the authors of those stories found such a representation fitting? Why do we agreeably acquiesce in the pictures and hence in the moral about the logical relations of belief to knowledge?

Behind representations of belief and knowledge as objects is the truism that both are *states*. What makes the analogies found in Plato and Ayer possible is that what is conceptually a state shares with objects (as contrasted with, say, numbers) the logical property of having a duration. What makes such pictures plausible is that believing and knowing *are* states. Because belief can persist it is possible and plausible to picture it as a statue, a dog, etc., and to further represent it as undergoing external modifications and so continuing to exist under the label "knowledge." If that is so (leaving confirmation to emerge later), the idea that cases of

⁵ A. J. Ayer, *The Problem of Knowledge* (Baltimore, 1956), pp. 16, 31.

passage show knowing to be a species of believing rests first of all on the truth that belief and knowledge are states.

Even so, that does not establish that the second is a continuation of the first. We pass from (the state of) sickness to (that of) health, yet that does not show health to be a variety of sickness. Some further requirements are best expressed by A. M. MacIver and Jonathan Harrison.

But if knowledge and belief are quite different in kind, it is oddly difficult ever to tell which is which.⁶

Being quite sure, alas, feels just like knowing, and the knower or believer cannot by introspection tell which is which. Furthermore, we behave with regard to the propositions we are certain of just as we behave with regard to the propositions we know. . . . Hence there must be an emotional and/or behavioral core, which knowing and believing have in common. He who knows, as well as he who believes, must believe.⁷

What is explicitly supplied here is the claim that knowing and believing are, as states, the same. This indistinguishability of one's state, i.e. of one's emotional and/or behavioral condition, is said to account for failures to tell whether we know or believe something.⁸ (It is not, of course, intended that *every* belief is indistinguishable from knowing.)

This claimed indistinguishability of condition is still not sufficient to reach the conclusion, as Harrison does, that knowing is believing. Even if one's emotional and/or behavioral condition is the same in both, even if there is a "common core," it would not follow that knowing is *believing* unless belief is taken to be that which is in common. Hence the argument requires, along with the indistinguishability claim, the implicit identification of belief with the common condition.

Notice that the argument being considered can function with any of the standard philosophical accounts of the nature of belief. Harrison uses the phrase "emotional and/or behavioral" to note that the indistinguishability doctrine is neutral between accounts of belief as a set of feelings and accounts of it as a pattern of behavior. Even more: it is neutral between such traditional construals and more recent dispositional interpretations. If belief

is taken to be a disposition to feel and/or behave in certain ways, then what is claimed is that being so disposed is a condition in which one finds oneself and hence one who knows has the same disposition. If in what follows these distinctions are blurred, it must not be thought that that has any effect on the proceedings.

What gives rise to this idea that "There is no difference, on the subjective side, between knowing and believing."⁹ The idea derives from a type of situation, the chief features of which are most vividly revealed in an example like the following. Suppose a man puts some cookies on a table; and that an hour later someone makes off with them; and another hour later the man is asked "Do you know where the cookies are?" and, having no reason to think they might have been taken, replies "Yes—on the table." It takes little reflection to see that, though the man knew where the cookies were the first hour and did not know the second hour, his feelings and behavior were (or would be) the same throughout the two hours.

More is needed to make out that it is knowing and believing which overlap here—nothing has yet been said about belief. MacIver follows his previously cited assertion of indistinguishability with a reference to such situations and concludes with the requisite claim that the second hour is a period of belief.

It is impossible to know what is not the case. But people sometimes claim to know what even they themselves later allow not to have been the case. It follows that they did not know, but only believed, what they falsely believed themselves to know.¹⁰

What is exhibited here is that same fascination with the area where believing and knowing meet. Only now the interest has shifted from the passage from belief to knowledge to a concern with what happens when a claim to know does not hold up. At its most dramatic, as in the example, what we are said to have is a reverse passage from knowledge to belief.

With these elements the description is complete. A recapitulation is in order. The basic form of the passage case argument is "The passage from belief

⁶ A. M. MacIver, "Knowledge," *Proceedings of the Aristotelian Society, Supp. Vol.* 23 (1953); reprinted in part in *Knowing: Essays in the Analysis of Knowledge*, ed. by Michael Roth and Leon Galis (New York, 1970), p. 6.

⁷ Jonathan Harrison, "Does Knowing Imply Believing?", *Philosophical Quarterly*, vol. 13 (1963); reprinted in Roth and Galis, *op. cit.*, p. 156.

⁸ There is a further assumption here—that it follows from the truth that belief and knowledge are states that "being a state" means "being in a certain emotional and/or behavioral condition." See note 18 for criticism.

⁹ Ayer, *op. cit.*, p. 9.

¹⁰ Though incomplete, the fullest articulation of this argument that I have found in print is by Roth and Galis, *op. cit.*, Introduction, p. 8. It is frequently suggested in discussion, most notably by Noel Fleming.

to knowledge shows that knowledge is (true) belief plus. . . ." Finding the additional thoughts which held together than *non sequitor* required extensive detective work. Those background ideas have been discovered to be: (1) the truism that belief and knowledge are states; (2) the thesis that as states, i.e. as the actual or potential feelings and/or behavior of a person, they are the same; that thesis is supported by (3) an appeal to situations in which a person who says he knows, does not know and (4) a claim that in such situations the person (only) believed where (5) the belief is identified with the inner and outer condition of the person. Given these elements, one who passes from belief to knowledge on a given point will remain in the same state, that of believing, but in having satisfied further essential conditions will also know. So we see that knowing entails believing.

The second major source of the thesis that knowing is believing—call it the "subjunctive argument" for reasons to be apparent—draws upon the same group of ideas found in the argument from passage cases. Using again the example of the cookies, this second argument can be succinctly stated. "As the cookies were not on the table the second hour, the man did not know but only believed they were there. But if they had been there, as they were the first hour, he would have known. Since the only difference between his knowing and his (only) believing is the cookies being on the table or not, it follows that he would have been in that state of belief even if they had been there, even if he had known. So it is clear that knowing is a special case of believing."

The chief means of support for the entailment thesis have been provided. It is time to begin the response to them. Consider that step of the arguments in which it is said that, in cases where someone claims to know but does not know, it follows that the person (only) believes. Does that follow? Is that a faithful account of what must be said (of what must be *thought*), when a claim to know goes awry?

Let us see how we can talk when the man answers "Do you know?" with "Yes—the cookies are on the table" though they are not. It turns out, even to the man's satisfaction that the cookies are not there. Must he sink back into the language of belief? No: a possible response is "Well, they *were* there," perhaps adding "I (recall having) put them there." The person may persist in that, trying to think what might have happened to them. In the third-person we may go along, not shifting into

belief talk. "John said the cookies were on the table but they weren't—someone must have taken them." Suppose it is learned that that is what happened. The man's response is "I *knew* they were there," "I *knew* I'd put them there." So we talk in the third-person too: "John *knew* he'd put them on the table but they were gone when we looked."

There are familiar cases then where, when talk of knowing has to be surrendered, there is yet no redescription of the case in terms of belief. Contrary to the allegation, it is not necessary that we understand failed claims to know as being cases of belief.

If belief talk is not a necessary consequence of a claim to know being mistaken, it seems that when (and if) it does become appropriate to talk so, there must be some special conditions for doing so. To learn about these, let us examine some variations on the example. After having said that yes, he knows where the cookies are, on the table, and after finding that they are not there, other outcomes are possible. The first of these still does not result in the use of belief language. Our man may say, rather than "I put them there," "They're not! Aunt Bessie told me that she had left them there." Again the third-party may fall in line: "We don't know where the cookies are. Aunt Bessie told John she had put them on the table, but they're not there." This variant overlaps the original in the following way. Having learned the facts to be otherwise, the man's aim becomes that of vindicating himself as the author of the claim to know by exhibiting the nature of his entitlement to have said he knew. The lesson seems to be that when a claim to know goes wrong, our actions and talk have the object of protecting one's authority as a source of saying how things are.

Consider another variation in which there seemingly occurs the sought for resort to belief language. The man has said he knows but the cookies are not there. A defense is appropriate—but it is not possible. For he does not *distinctly* recall having put the cookies on the table. He *meant* to, but he was busy, preoccupied with stock market. Maybe he hadn't put them there after all. Still preoccupied, he had answered "Yes, I know" rather than using the more accurate "No, I don't know, but I think that. . . ." Now there is nothing to say except "Damn! I *thought* I had put them there." Such a case suggests that we fall back on "I thought" precisely where no defense of our original claim is available. It would be a mistake, however, to think that with the invocation of "I thought" we have left the arena where our responsibility for

having said we knew is a tissue. A vindication of the words "I know" may not be possible, yet we need to assure others that when we claimed to know we were not merely guessing or intending to overstate the case or even lying, joking, etc. Saying "I *thought* so" does just that. By saying "I *thought*" our cookie man admits to having spoken without due consideration, but otherwise responsibly, sincerely.

What we are *not* doing in saying "I *thought*" is saying "I didn't know it (the facts were not right), so I only had the belief that it was so." That is, even with "I *thought*," we still have not found that belief talk required by the arguments for the entailment thesis. "I *thought*" here is not a past tense of "I believe," it is not the assertion of a prior belief. "I *think*" and "I *believe*" are often equivalent; e.g. "I *think* it's over there," "I *believe* it's over there." Moreover they are interchangeable as prefates to "I *know*"; e.g. "I *think* I know where to find it," "I *believe* I know where to. . . ." Philosophers have been led from those facts into taking it that in the cases of interest "I *thought* I knew" and "I *thought*" amount to "I *believe*." But my ear tells me that we never say "I *believed*" in such circumstances. A man responds confidently "Yes" to his wife's "Now you're sure you know where they live?" and then cannot locate the house. He may say "I *thought* I knew where it was, I *thought* it was right here." I cannot hear him saying "I *believed* I knew where it was, I *believed* it was here."

The above line of criticism has held that, given a failed claim to know, a redescription of the case in belief language is not necessary and that even when putative belief talk does occur it is not the genuine article. Let us now turn to a different objection.

The subjunctive argument goes "As the cookies were not on the table the second hour, the man did not know but only believed they were there. But if they had been there, as they were the first hour, he would have known. Since the only difference between his knowing and his believing is the differing location of the cookies, it follows he would have been in that same state of belief even if he had known." What does this argument rest upon? Obviously it depends upon the claim that the second hour is a period of belief. Setting aside, for now, the previous criticism of that, let it be granted. Next, the argument depends upon what is indeed exhibited by the example, the continuity of the man's condition through both hours. While these items are sufficient for the thesis that knowing and

believing are, in these cases, subjectively the same, they do not prove that knowing is a form of believing. Pointing out how the man's condition is the same when he knows and when he believes but does not know would not enable one to conclude that he believed when he knew unless it is taken that the believing is just that which is indistinguishable in the two periods. The belief must be identified with the man's condition.

To discover the crisis lurking here consider the following. The phrase "did not know but only believed" occurs importantly in comments on the example. It is clear that the two hours are different in that the man knew and then did not know where the cookies were. While the official lesson of the example is the sameness of the states of one who knows and one who believes, it is much more certainly a lesson in the sameness of condition of one who knows and one who does not know. That lesson suggests a parody of the subjunctive argument. "Since the only difference between his knowing and his not knowing was whether the cookies were on the table or not, it follows that he would have been in that same state of not knowing even if he had known. So we see that knowing is a species of ignorance."

The parody turns, analogously to the original, on identifying the man's not knowing with that condition which persists for the two hours. That is illegitimate and a defender of the subjunctive argument will say so. But such a defender must maintain that he is entitled to the identification of belief with the man's condition, while denying the propriety of the parallel identification. How is the distinction to be made? For it is a clear case of not knowing, yet *that* is how the man feels and acts; *that* is what his not knowing amounts to, yet it is precisely what his (claimed) believing amounts to. The only plausible way of rejecting the second identification is to say that those feelings, etc. are not typical of not knowing something; what is typical is that emotional and behavioral state which goes with a person's saying "I don't know." That line has the virtue of truth. It brings out why we resist calling the man's condition "the state of ignorance." Despite its truth, as a defense it concedes too much. For, *mutatis mutandis*, our man's feelings and behavior are just as untypical of believing; what is characteristic of that will be found where a person says "I believe." To reject the one identification on those grounds (and nothing else is plausible) is to reject the other identification also.

The wisest course is to admit that both identifications are improper. Upon realizing that the second hour is a period of not knowing, we rightly regard it as accidental and not essential that the man is in *that* condition though he does not know. It would be equally accidental, if we ever talk about belief here, that he is in *that* condition and yet believes. The accident was the unnoticed and unexpected removal of the cookies. The analogy is this: a man is demoted from sergeant to corporal unbeknownst to him. So he goes about in the same old way, acting, thinking and feeling like a sergeant. It would be totally wrong to identify his being a corporal with his present condition and in consequence to hold that being a sergeant is also being a corporal. It is not less misguided to identify believing with the condition of a man who does not realize that he has been epistemically demoted and in consequence to hold that knowing is believing.

Both the subjunctive and the passage case arguments depend upon the indistinguishability thesis. That thesis probably does not survive the earlier criticism of the claim that we fall back on belief language when a knowledge claim fails. However, even if there can be a subjective sameness of knowers and believers, the identification of the belief with that which is indistinguishable is illegitimate. Lacking that requirement, *both* arguments fail as proofs that knowing is believing.

The thesis that knowing is believing is not established by those considerations which have been historically most prominent and which are logically the most compelling. It is time to take up the task of arguing that knowing is necessarily not believing.

Those arguments for the entailment thesis pivot on the presumption that the belief is identical with that which is indistinguishable through the two periods. What leads to that identification of a belief with a person's condition? It is the general idea that belief *is* (and is nothing else but) a set of feelings and/or behavior, actual or potential. Associated with that is the idea that knowledge too *is* (though in part only) such a collocation. In short, the best arguments for the entailment thesis rest on a certain view of what knowing and believing are, of their nature or essence. The accounts in question identify belief and, in part, knowledge

with the subjective and behavioral phenomena of believing and knowing.

Despite such accounts being widely accepted, they are not acceptable. The idea that belief is a pattern of feelings and/or behavior gets a start because it is supposed that to believe such-and-such is to act and feel in a certain way. That supposition has been demolished by Peter Geach. (While Geach speaks only of behavior, the same points apply to feelings.)

But is there in fact any behavior characteristic of a given belief? Can action be described as 'acting as if you held such-and-such a belief' unless we take for granted, or are somehow specially informed about, the needs and wants of the agent? In Ryle's example this information is smuggled in by his speaking of a *gardener's* rain-expecting behavior (and tacitly assuming that the gardener is not e.g. a discontented or corrupt servant who wants the garden to be ruined.) When Dr. Johnson did penance in Uxeter market-place, he may have begun by standing around bare-headed until the threatened shower should fall; this would not be recognizable as rain-expecting behavior without a knowledge of Johnson's wish to do penance.¹¹

Geach makes the same kind of point made by Socrates to Cephalus in the *Republic*: right behavior can be either returning something entrusted to one or not returning it. So too belief that it will rain can be expressed either by going out with a hat or by not taking a hat; it can be expressed by feeling gleeful or by feeling depressed. There is no such thing as a pattern of behavior and feeling which is in itself that of believing it will rain. And so for other beliefs.

There is an objection to this. "Geach's point establishes only that there is no behavior (and feelings) characteristic of a *given* belief. The issue, however, is belief itself, whether there is any subjective and/or behavioral condition which constitutes belief as such. There must be such a topic-neutral condition—what else could belief be?"

The favorite candidate for such a topic-neutral condition has been confidence, comprising feelings of confidence and a confident cast to one's relevant behavior. Other specifications are possible, e.g. Hume's "superior *force*, or *vivacity*, or *solidity*, or *firmness*, or *steadiness*" of "ideas assented to" as contrasted with "a fictitious idea";¹² or being prepared to act (in whatever way is relevant to the

¹¹ Peter Geach, *Mental Acts* (London, 1957), p. 8.

¹² Hume, *A Treatise of Human Nature*, Appendix, Selby-Bigge edition (Oxford, 1967), p. 629.

given belief and to one's other beliefs and attitudes.) It is, however, beside the point to worry about what content to give to that topic-neutral condition of a believer, for to take any of those options is to mistake the cart for the horse.

Socrates in the *Euthyphro* poses the question "Do the gods approve an act because it is holy or does the nature of holiness consist in the gods' approval?"¹³ The parallel question about belief is overdue: is belief nothing else but acting and/or feeling *so* (to be filled in as you please) or does a person feel and act *so* because he believes? Surely the latter—our relevant subjective and behavioral condition (whatever it may be) is a consequence of believing. One is prepared to act—because, *inter alia*, one believes; this proposition is more "vivacious" than that—because this is believed but not that; why is John so confident?—because he believes. . . . The idea that belief is some inner and/or outer condition of people who believe fails because belief is explanatory of such a condition rather than identical with it.

All this applies to knowledge also. There is no pattern of behavior or feelings characteristic of a given piece of knowledge—how things are with one given the knowledge, e.g., that it will rain depends upon one's intentions, attitudes, etc. Any topic-neutral state of one who knows (claims to know) arises from the person's taking it that he knows—e.g. one is prepared to act, or feels confident, *because* one is willing to claim knowledge. Like belief, knowledge is the horse, our feelings and behavior the cart.¹⁴

I have lumped all the standard positions about the nature of (the proper "analysis" of) belief and knowledge together. The usual format for discussions is to say that knowledge is (something like) true justified belief. Then the disputes arise as to whether belief is an occurrence or a disposition, with the further issues about behavior and inner states cutting across those. H. H. Price, in *Belief*, attempts a thorough discussion of the topic—yet he says "There are two very different ways of analyzing belief" (the occurrential and the dispositional), and he clearly means and examines *only* two.¹⁵ In the preceding, all those recognized positions are

treated as amounting to the same thing, as a wrong-headed *kind* of answer to "What is belief (knowledge)?" What other kind of answer is possible, what other kind of thing could they be?

My aim will not be to attempt to answer "What are knowledge and belief?" but will be the more limited project of showing that knowing is not believing. Consequently, I shall be citing only some of what goes into an answer to those questions, shall sketch what *sort* of thing knowledge and belief are, and shall develop the material just far enough to achieve my aim. Nothing is claimed about the relative importance of items cited, either among themselves or to what is omitted, or about how such items would fit together in an account of the nature of knowledge and belief. Lastly, there is nothing mysterious about the elements I shall cite—most have appeared (without dispute) in the literature. The trouble is that the proper moral has never been drawn *pointedly*. That is the aim here.

First, (some of) the data must be presented, though with no order. It makes sense to say both "knows that" and "believes that"; it also makes sense to say "knows why, how, where, what, etc."; but "believes how, why, etc." are senseless. (One cannot ignore the grammatical objects other than "knows that" for every "knows that" is also a "knows how or why or . . ." and so too conversely.) We can speak of "beliefs" but not of "knowledges." It can be said of a belief that it is stupid or sensible, obstinate or hesitant, well- or ill-founded, child-like or mature; none of those things make sense about knowledge. We ask: "How do you know?" but "Why do you believe?"; "How intimately do you know?" but not "How intimately do you believe?"; "What makes him believe?" but not "What makes him know?" One may say "I cannot help believing" but "I cannot help knowing." Finally, that old standby "He truly/falsely believes that . . ." makes sense though "He truly/falsely knows that . . ." does not. The above is a brief selection of the differences between knowing and believing with respect to what it makes sense to say.¹⁶ Since such differences are familiar, to go on would be to multiply instances beyond necessity.

¹³ *Euthyphro* 10a.

¹⁴ All of this applies also to attempts to replace feelings and behavior in the formula with brain-states; one's brain is in state X because one believes or knows and so belief and knowledge cannot be identical with some brain-state.

¹⁵ H. H. Price, *Belief* (London, 1969), p. 19.

¹⁶ For these and more see: Gilbert Ryle, *The Concept of Mind* (London, 1949), pp. 133-134; J. L. Austin, "Other Minds," *Philosophical Papers* (Oxford, 1961), p. 46, footnote p. 48, and *passim*; Alan R. White, "On Claiming to Know," *The Philosophical Review*, vol. 66 (1957), reprinted in A. Phillips Griffiths (ed.), *Knowledge and Belief* (Oxford, 1967), p. 106.

The problem is how to take those differences, what it is that they show. Are they merely curiosities? Far from it. Such remarks as "knowledge cannot be sensible or stupid" tell us (something) about the nature of knowledge. To know *what belief is* (to have the concept of belief) is, among many other things, to know that it makes sense to ask "Why do you believe?"; but not to ask "How do you believe?" That we cannot speak of "knowledge" is a piece of what Wittgenstein called the "grammar" of a term and he added "*Essence is expressed by grammar*" and "Grammar tells what kind of object anything is."¹⁷ As the remarks constituting the data reveal extensive and significant grammatical differences (differences in what it makes sense to say) between belief and knowledge, it follows that they are two different kinds of "object."¹⁸ Knowledge and belief form two "different categories of being, different sub-languages within a language."¹⁹ Think of it this way. Hitler was not a very, very naughty man.²⁰ While we may expect that evil is an extreme form of naughtiness, that is not so. Between being naughty and being evil there is a *categorical* shift in the ways we talk and think, a point at which (in Hegelian terms) a quantitative difference becomes qualitative. So too knowing is not a (true) belief which is very, very well supported. At some point to acquire more "evidence" is to move into a radically different web of language, with consequent alterations in behavior and feelings.

Others have held that knowledge and belief constitute different categories and did so on grounds of grammatical differences. While Plato, in the *Meno* and *Theatetus*, is the source of the idea that knowing is believing, the Plato of the *Republic* (477b-478b) is the ancestor of the present view. He held that knowing and believing are "different things/faculties;" on the ground that one is fallible, the other infallible—that being a misguided characterization of such grammatical differences as that between the propriety of "He mistakenly believes" and the impropriety of "He mistakenly

knows." Plato concluded from that difference (again erroneously because misled by talk of "faculties" and a consequent comparison of knowing and believing with seeing and hearing) that the objects of knowledge and belief cannot be the same. This kind of position has been revived recently, stripped of the mistaken features of Plato's account. Ryle: "'know' and 'believe' . . . are . . . verbs of quite disparate types"; Austin: "Saying 'I know' is taking a new plunge. But it is *not* saying 'I have performed a specially striking feat of cognition. superior, in the same scale as believing and being sure, even to being merely quite sure': for there is nothing in that scale superior to being quite sure;" Wittgenstein: "Knowledge and certainty belong to different categories."²¹

Those prior claims to a difference of logical type or category or "scale" between knowledge and belief have not affected the subsequent discussion of whether the one entails the other. That is because their implications for the entailment problem were not worked out. Those implications need to be even more pointedly marked. We saw earlier how the relation of knowing to believing was offered *pictorially* in connection with the passage case argument (belief as a statue, a dog). To make the relation of knowing to believing argued here perspicuous, a representation superior to those seems appropriate. (Wittgenstein's warning "A picture held us captive"²² should not lead us to eschew pictures—the point is to avoid captivity.)

What would be a fitting model of what it is to pass from believing something to knowing that same thing? Here are two suggestions which incorporate the results of the preceding investigation. Passing from belief to knowledge with respect to the same proposition is like passing from playing volleyball to playing soccer with the same ball. Again: it is like spending the first part of the evening playing poker, then using the same cards to play bridge. The cards are the same, but the framework in which they are deployed is vastly

¹⁷ Ludwig Wittgenstein, *Philosophical Investigations* (Oxford, 1958), #371 and #373.

¹⁸ I said that it is true that knowledge and belief are states. Readers may be puzzled as to how it can be affirmed that they are states and yet denied that they are a set of feelings and behavior. "Is a state," where used so that it is true of knowledge and belief, is a grammatical remark, material mode for, e.g., "He has known/believed for years," "He still knows/believes." In defending the entailment thesis, the inference from "belief and knowledge are states" to "belief and knowledge are an inner and/or outer state (condition) of the person" is fallacious, involving an equivocation on "is a state."

¹⁹ John Wisdom, "Philosophical Perplexity," *Philosophy and Psychoanalysis* (Oxford, 1964), p. 37.

²⁰ I owe this example to J. K. Stephens.

²¹ Ryle, *op. cit.*, p. 133; Austin, *op. cit.*, p. 67; Ludwig Wittgenstein, *On Certainty* (Oxford, 1969), #308.

²² Wittgenstein, *Philosophical Investigations, op. cit.*, #115.

different and so too (in consequence) is our condition, whether as players or as spectators.²³

Those models are helpful in preventing misconceptions of the force of the grammatical differences between knowing and believing. Among defenders of the entailment thesis, Chisholm has most fully recognized the range and significance of those differences.

We must not think of knowing as being, in any sense, a 'species' of believing. . . . A man can be said to believe firmly, or reluctantly, or hesitatingly, but no one can be said to *know* firmly, or reluctantly. . . . Austin has noted that, although we may ask, 'How do you know?' and 'Why do you believe?' we may not ask, 'Why do you know?' or 'How do you believe?' The relation of knowing to believing . . . is not that of falcon to bird or airedale to dog.²⁴

Still, having gone this far, Chisholm believes he can draw a distinction between knowledge being a kind of belief and knowledge entailing belief without being a species of it and thereby save the entailment thesis.

[the relation] is more like that of arriving to traveling. *Arriving* entails *traveling*—a man cannot arrive unless he has traveled—but arriving is not a species of traveling.

That analogy was undoubtedly suggested by passage cases ("arrives at knowledge"). But it is so

far from upholding his view that it belongs to mine. Arriving does not entail traveling—it entails *having* traveled. When one has arrived, one is no longer traveling. Similarly, when one does come to know what was previously believed, the believing, like the traveling, is over.

More must be said. Chisholm's analogy goes against him, but its failure does not show there to be no entailment. It remains possible that knowing entails believing even though it cannot be a species of belief. What plausibility is there, however, to the idea that when a person knows, i.e. is embedded in one system of talk, another, very different system, *bearing on the same proposition*, must also be applicable? It is here that those models are helpful. What they encapsulate are such things as: when someone knows something, e.g. where Godot is, that person cannot try to find out, e.g. where Godot is; while if someone believes something, that further inquiry bearing on what is believed is possible. Compare: "I think he's not around but I'm still trying to make sure" and "I know he's gone but I'm attempting to find out." A man who knows cannot do what someone who believes can do. Similarly, someone playing poker cannot trump, while that is possible in bridge. Asserting that knowing entails believing is like saying that to be playing poker is also and necessarily to be playing bridge.²⁵

California State University, Fullerton

Received November 25, 1975

²³ The feature of the analogy that I find most unhappy is one that shall probably commend it to many. The grammar of a concept is not a system of rules for the use of a term. For the best criticism of this fashionable identification, see Hubert Schwyzer, "Rules and Practices," *The Philosophical Review*, vol. 78 (1969), pp. 451-467.

²⁴ R. M. Chisholm, *Perceiving* (Ithaca, 1957), pp. 17-18.

²⁵ One should compare the oddity of "He knows it but he doesn't believe it" with the similar oddity of "They're playing bridge, but they're not playing poker." It goes without saying that if the first is so, the second is excluded—so it is most odd to say that the second is excluded.

VI. THE PRINCIPLE OF PATERNALISM

JOHN D. HODSON

THE classical treatment of the problem of paternalism was, of course, that presented by John Stuart Mill in *On Liberty*. While both the details of Mill's position and the merits of his arguments are matters which have generated much debate,¹ there seems to be no question that Mill's position includes (i) the claim that there is a class of persons or actions with respect to which paternalistic interventions can never be justified, and (ii) the claim that there is a class of persons, e.g., children, with respect to which paternalistic interventions can be justified. Mill seemed to be primarily interested in explaining and justifying the first of these claims, and so devoted little attention to the problem of distinguishing cases which fall under the first claim from cases which fall under the second. While Mill was no doubt justified in this focus of attention, he thus leaves some of the most difficult problems concerning paternalism unanswered.

Recent writers on paternalism seem to display a consensus of opinion on the question of the correctness of Mill's major claims: in general, paternalism is to be forbidden, but in certain kinds of cases it may be acceptable.² There is, however, no consensus on the question of just how the line between cases in which paternalism is acceptable and those in which it is not acceptable is to be drawn. Accordingly, the problem to be considered here is that of establishing criteria for identifying cases in which paternalism is justified. To do this, I shall propose necessary and sufficient conditions of justified paternalistic intervention; the adequacy of these conditions will be tested by reference to their implications for intuitions about paternalism which would be shared by Mill and those who agree with his major claims.

I

It will be useful to begin by specifying some of the cases about which there seems to be agreement that paternalistic interventions can be justified. Six reasonably distinct categories may readily be identified.

Ignorance. Cases involving potential harm of which the threatened person is unaware provide clear examples of paternalistic coercion widely thought to be justified. For instance, it is thought to be unobjectionable to forcibly prevent children from getting themselves into dangerous circumstances which they do not understand in a similar way, compulsory education of children is widely accepted partly because children are unlikely to know the extent to which they will need education in adult life. However, not all examples of this type involve children; if an adult is in a seriously dangerous situation and is unaware of it, it can also be permissible to intervene paternalistically. For instance, if a man unknowingly is standing in the path of a runaway truck, surely a passerby could be justified in pushing him out of danger.

Emotional stress. People sometimes decide to inflict harm on themselves when they are in periods of great emotional stress. If, for instance, a person attempts suicide while subject to extreme and unusual stresses, it seems that another party would be justified in intervening in at least some such cases.

Compulsion and Undue Influence. Self-harming behavior can also be caused, it seems, by action done under compulsion or undue influence. If self-harming behavior is due to, say, some irresistible psychological influence, or to threats of greater, other-imposed harm, many would hold that interference with the behavior could be justified even if the failure to interfere would harm no one other than the person interfered with.

¹ See, e.g., C. J. Friedrich (ed.), *Liberty* (New York, 1962), and Peter Radcliff (ed.), *Limits of Liberty* (Belmont, Calif., 1966).

² Recent discussions include Gerald Dworkin, "Paternalism" in *Morality and the Law*, ed. by Richard A. Wasserstrom (Belmont, Calif., 1971); Joel Feinberg, "Legal Paternalism," *Canadian Journal of Philosophy*, vol. 1 (1971), pp. 105-14; C. L. Ten, "Paternalism and Morality," *Ratio*, vol. 13 (1971), pp. 56-66; David B. Wexler, "Therapeutic Justice," *Minnesota Law Review*, vol. 57 (1972), pp. 289-338; and Jeffrie G. Murphy, "Incompetence and Paternalism," *Archiv für Rechts- und Sozialphilosophie*, vol. 60 (1974), pp. 465-486.

Mental Illness. Closely related to the previous category is that of persons labeled mentally ill or insane. When such persons behave in self-destructive ways, and when this self-destructiveness appears to be part of their affliction, paternalistic intervention again appears to be justifiable.

Nonrationality. Sometimes persons who are in states of impaired consciousness can be the subjects of apparently justified paternalism. A person who is, for instance, unconscious due to a blow received in an accident would be unable to decide for himself whether to do such things as go to a hospital; in such a case it hardly seems objectionable for someone else to take the unconscious person to a hospital.

Serious harm. Sometimes the mere fact that a person is about to cause himself serious harm may be taken as sufficient to justify paternalistic intervention even if the person is not known to fall into any of the above categories. If, for instance, there is no time to ascertain whether the person threatened with harm has chosen to be in that position, surely it could be justifiable to interfere with his action at least long enough to determine whether the person falls into any of the foregoing categories.

These categories represent various factors which seem to play a role in the justification of paternalism. It is worth noting, however, that the specific role played by each is not parallel throughout the six categories distinguished. The first five appear to represent states of mind under which a person may undertake possibly self-destructive action; by contrast, the last category—that of serious harm—does not represent a state of mind at all, but instead a possible consequence of action. Note also that the threat of harm not only represents a category which in itself can sometimes be sufficient for apparently justified paternalism, but also one which plays a role when other categories apply as well; that is, both the threat of harm *and* one of the states of mind are present in examples involving the first five categories. This suggests that the threat of harm may play a role in the justification of paternalism which is different from that played by the factors in the first five categories.

One of the problems of paternalism is that of organizing the factors represented by these categories in such a way as to provide some way of distinguishing cases in which paternalism is justified from cases in which it is not. This would involve showing when, why, and how these ele-

ments are relevant to paternalistic justification. Philosophers who have studied paternalism have proposed various principles designed to do this. There is fairly close agreement between the various proposed principles, and while it seems that each of them does *approximate* the sort of principle which captures common intuitions on this matter, I shall argue that each of them does have certain counterintuitive consequences.

II

Consider first the approaches suggested by Gerald Dworkin. He initially suggests as a general principle the following limitation on paternalistic interventions: "Paternalism is justified only to preserve a wider range of freedom for the individual in question."³ This principle certainly has an attractiveness, at least at first glance. However, it cannot be taken as an adequate principle of paternalism. For one thing, if it is taken literally it states only a necessary condition of justified paternalism; an adequate criterion would state both necessary and sufficient conditions. But even the necessary condition suggested is questionable, for there are intuitively acceptable cases of paternalism which would be ruled out by this principle. Suppose a child is about to touch an object which is hot enough to burn his or her hand, but that the child does not realize this. Physically restraining the child from touching the object hardly seems objectionable, but it appears to be ruled out by this principle, for having a burned hand is surely not a limitation on the child's freedom in any plausible sense. The protection of freedom is only one thing for which paternalism may be acceptable.

Another problem with Dworkin's principle can be seen if we take it as giving a sufficient condition of justified paternalism. Taking it in this way is not altogether unfair to Dworkin despite his wording of the principle, for his use of it in connection with contracts selling oneself into slavery suggests that he may have intended it as giving a sufficient condition. In this regard, he claims that "the main consideration for not allowing such a contract is the need to preserve the liberty of the person to make future choices."⁴ Since no harm other than loss of liberty need come from contracting oneself into slavery, the loss of liberty must count as a sufficient condition if all slavery contracts are to be ruled out. But Dworkin's principle is not acceptable

³ "Paternalism," *op. cit.*, p. 118.

⁴ *Ibid.*

when interpreted in this way either because if the preservation of a wider range of freedom were a sufficient condition of justified intervention, not only would slavery contracts be ruled out, but so too would virtually any contract in which a person agreed to commitment of his future services. As Joel Feinberg has pointed out, employment contracts often involve limitations on the liberty of both employer and employee. Moreover, he cites cases such as contracts "in restraint of trade" in which courts have supported the possibility of contracting away extensive liberties in exchange for some benefit.⁵ Thus, preservation of a wider range of freedom cannot be the crucial question.

Dworkin also suggests another criterion which he calls that of "future-oriented consent."⁶ The primary example of this is the case of parental paternalism which, Dworkin suggests, "may be thought of as a wager by the parent on the child's subsequent recognition of the wisdom of the restrictions."⁷ However, as Jeffrie G. Murphy has argued,⁸ there are at least two problems with this model. First, there is the danger that the consent to paternalism may be artificially manufactured: the paternalistic intervention could include the attempt to condition the person to accept the intervention. Second, there is the question of what should count as future consent. The danger here is that the failure to agree to the paternalistic intervention would never be accepted as the decision to be taken seriously. Beyond this, there is a question about the range of application of this model. If the justifiability of paternalistic interventions is really to depend on future consent, then there is no way of knowing at the time of the intervention whether it is justified. Thus, the future consent model is not the answer.

The approach of Dworkin's on which he seems to put greatest emphasis, however, is that which appeals to what would be chosen by fully rational individuals. He claims that we "may argue for and against proposed paternalistic measures in terms of what fully rational individuals would accept as forms of protection."⁹ This approach is clearly quite similar to the Rawlsian method of argumenta-

tion in terms of what would be chosen by parties in the original position.¹⁰ The main difficulty with it, however, is that it involves a serious departure from the position of Mill and those who agree with his main claims. For Dworkin's position essentially involves the substitution of an abstract "rational will" for the will of an individual; such substitutions, as is argued below (section IV), result in a paternalism which is too strong to capture the intuitions with which we are here concerned. Thus, despite the suggestiveness of much of what Dworkin has to say, he has not provided a generally applicable principle for identifying cases of justified paternalism.

Consider next C. L. Ten's discussion of paternalism. He suggests the "limitation of paternalism to cases where either one of the following two features are present: (a) the agent's decision is clearly and seriously impaired . . . ; and (b) the harm inflicted on the agent is of a severe and permanent type."¹¹ As stated by Ten, this principle appears to give merely a necessary condition of justified paternalism; if this is what is intended, the principle is inadequate because it fails to indicate what is sufficient for intervention. However, Ten's discussion suggests that he intends the principle to provide both necessary and sufficient conditions. But on this interpretation the principle still involves serious problems.

One problem is that it allows paternalism when the agent's decision is impaired without imposing any further restrictions. But this is surely not enough. Surely paternalism should be allowed only when the impairment is likely to have undesirable consequences for the agent. If nothing undesirable will result, or perhaps if the agent will learn through his mistake, it seems doubtful that intervention should be allowed. Perhaps Ten meant these considerations to have been accounted for in the qualification that the decision be *clearly and seriously* impaired; but if so this should be further explicated.

A second problem with Ten's principle is that it allows intervention when there is severe and permanent harm, and without further qualification. This sounds as though it would prohibit anyone

⁵ Feinberg, "Legal Paternalism," *op. cit.*, pp. 121-122.

⁶ "Paternalism," *op. cit.*, p. 119.

⁷ *Ibid.*

⁸ "Incompetence and Paternalism," *op. cit.*, p. 482n.

⁹ "Paternalism," *op. cit.*, p. 120.

¹⁰ John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971). The Rawlsian approach is explicitly adopted and developed in Murphy, "Incompetence and Paternalism," *op. cit.*

¹¹ "Paternalism and Morality," *op. cit.*, p. 65.

from taking his own life, for example. Death is surely a severe and permanent harm. But to accept this is to seriously depart from the position of Mill on paternalism, for Mill was concerned to argue that persons should be allowed full control over actions which do not seriously affect others; provided that the person is in full command of what he or she is doing, he/she should not be interfered with provided that others' interests are not harmed. Since Ten's principle fails to impose such restrictions, it cannot be the one being sought.

It may appear that Ten's problems could be avoided if, instead of requiring that only one of his two conditions be fulfilled, he were to require that they both be fulfilled. This would be to require both impairment of decision and serious harm, thus ruling out paternalism in cases like those used against the principle as originally stated. But even with this modification, the principle is not acceptable, for paternalistic interventions may be justifiable even when there is not a threat of severe and permanent harm. A child about to burn his finger is not in danger of such harm, and yet intervention seems unobjectionable. Thus, this modification of Ten's principle is another example of a principle which is perhaps close to what we want, but which fails to accurately capture common intuitions on paternalism.

Let us next consider the paternalistic principle proposed by Joel Feinberg. Feinberg suggests what he calls "the standard of voluntariness": "the state has the right to prevent self-regarding harmful conduct only when it is substantially nonvoluntary or when temporary intervention is necessary to establish whether it is voluntary or not."¹² Feinberg, as indicated by the principle, is primarily concerned with *legal* paternalism, but it does not seem implausible to assume that the standard of voluntariness might apply to paternalistic interventions undertaken by private individuals as well. In any case, this principle, like the others, has certain problems. One of them is the fact that it states only a *necessary* condition of justified intervention. As we have seen, unless something undesirable for the person subjected to paternalism is likely to result in the absence of intervention, there is no reason to suppose that intervention is justified. Thus, one would like to know what else is necessary, what is sufficient, for justified paternalism. Feinberg's principle is of no help here.

A second difficulty with the standard of voluntariness is its dependence on the concept of voluntary action. One reason why this is a problem is the fact that Feinberg nowhere provides a general account of voluntary action; he merely provides some examples of involuntary actions without considering what makes them so. The application of the standard surely requires some further understanding of this concept. A deeper problem with Feinberg's use of the concept of voluntary action is the fact that it may rule out cases which seem to be instances of justified paternalism. One situation in which paternalism appears justified is that in which a person attempts to inflict serious harm on himself while under severe emotional stress. To restrain someone attempting to commit suicide, for example, sometimes seems acceptable. But it does not seem plausible to suppose that a person is not acting *voluntarily* because he is acting under emotional stress; the standard of voluntariness would therefore rule out such interventions. Hence, either the standard of voluntariness must be given up or the concept of voluntariness must be extended to cover cases of emotional stress. In either case, further work is required, and so the standard of voluntariness is not the whole story.

The most recent discussion of paternalism to be considered here is that of Jeffrie G. Murphy. Murphy suggests the following principle of paternalism: "When a person is incompetent in any of the ways specified above, and if his incompetence is likely to result in major and not easily reversible harm to him, then paternalistic intervention is justified provided that it is carefully specified, limited, controlled, and explicitly tailored to the kind of incompetence manifested."¹³ The forms of incompetence Murphy specifies include cases in which a person is "ignorant, compulsive, or devoid of reason"¹⁴ with respect to the area involved. Murphy is primarily concerned with legal paternalism, and in that area his principle appears quite useful. However, it does not provide a *general* criterion for the justification of paternalism, for, I shall argue, there are clear cases of acceptable paternalistic intervention not allowed on this principle.

There is a difficulty in assessing Murphy's principle due to the fact that the role of the major-and-not-easily-reversible-harm requirement

¹² "Legal Paternalism," *op. cit.*, p. 113.

¹³ "Incompetence and Paternalism," *op. cit.*, p. 479.

¹⁴ *Ibid.*, p. 468.

is unclear. In the quotation above this requirement appears to be one element in what will count as a sufficient condition of justified paternalism, but it is not claimed to be a necessary condition, for the principle does not rule out the possibility that there might be other sufficient conditions not including the harm requirement. However, in his discussion of the principle, Murphy makes it sound as though this requirement is intended as a necessary condition: "Liberty is too important a value, too primary a good, to be interfered with except in the clearest cases of major harm."¹⁵ However, since I shall argue that there are acceptable cases of paternalism where there is no threat of major harm, the principle can be criticized whichever way it is interpreted. For if there can be justified paternalism without the threat of serious harm, then the harm requirement is not a necessary condition of such intervention nor is it part of the only sufficient condition. So either the principle incorrectly excludes such cases or it fails to capture some conditions under which paternalism is justified. But let us consider an instance of this type of case.

As indicated, the sort of case I have in mind here is that in which there is no threat of major and not easily reversible harm, but in which paternalism nevertheless seems clearly justified. Consider the case of John Doe who is about to take a drink of what he believes to be a glass of wine. Mary Doe, however, knows that the glass does not contain wine at all, but instead a harmless but foul-tasting liquid which only looks and smells like wine. It seems quite clear that Mary, knowing that John is acting on a false belief, would be quite justified in staying John's hand although he is in no danger of undergoing major and not easily reversible harm.¹⁶ Similarly, paternalistically preventing minor injuries to others is surely sometimes justifiable. Thus, Murphy's principle also fails to provide the criteria being sought.

On the basis of the foregoing, I conclude that no criterion of justified paternalism has been proposed which provides necessary and sufficient conditions of justified paternalism in such a way as to capture fairly common intuitions about when such action is justified. I should like now to propose such a criterion. By way of intuitive justification, I shall argue that the principle I shall propose not only correctly distinguishes cases in which paternalism seems clearly justified from cases in which it does

not, but also that the principle provides a highly plausible way of systematizing the various factors relevant to the determination of whether paternalism is justified in a given case.

III

I propose the following principle, which I shall call "the Principle of Paternalism," as a principle for the determination of whether paternalism is justified: Paternalistic interventions are justified if and only if (i) there is good evidence that the decisions with respect to which the person is to be coerced are encumbered, and (ii) there is good evidence that this person's decisions would be supportive of the paternalistic intervention if they were not encumbered. Let us consider some of the important concepts included in this principle.

Good evidence. The Principle of Paternalism relies on the same concept of good evidence as that used in any other area in which rational support for assertions is required. In general, evidence for a claim may be considered good evidence when it is factually correct and it serves as the premiss or premisses of a nonfallacious inductive or deductive argument in which the claim serves as the conclusion. Thus, if the argument is deductive, good evidence is provided only if the argument is valid and the premisses are true. If the argument is inductive and its premisses are true, the evidence provided is only as good as the strength of the inductive argument. Since the strength of inductive arguments varies in degrees, the strength of the evidence for an inductively supported conclusion can also vary in degree. Also relevant to the value of an inductive argument is the question of whether there is also evidence which supports claims incompatible with the conclusion of the argument. Evidence would not count as good evidence for a claim if an incompatible conclusion is supported by a stronger argument. These features of inductive argument are especially important for paternalism, for much of the crucial evidence in its justification is empirical and thus depends ultimately on inductive reasoning.

Encumbered decisions. The Principle of Paternalism allows paternalistic intervention only when there is good evidence that a person's decisions are encumbered. But what is it to say that someone's decision is *encumbered*? I shall say that a person's

¹⁵ *Ibid.*, p. 483.

¹⁶ Cf. Murphy's example of the man unknowingly choosing poisoned items from a buffet table. *Ibid.*, p. 469.

decisions are encumbered whenever those decisions are made in circumstances which are known to affect decision-making in such a way that the person making the decisions sometimes comes to believe that the decisions were mistaken or unfortunate. Thus, if a certain state, e.g., ignorance, is known to cause people to make decisions they later come to regret, that state constitutes an encumbrance of decisions. This means that the question of whether a given condition constitutes an encumbrance often depends on empirical evidence. Ignorance, for example, counts as an encumbrance because it is known that people who make decisions while lacking relevant information often come to regret their decisions. Emotional stress counts as an encumbrance for the same reason: people often come to regret decisions they have made and actions they have taken while under such stress. The set of circumstances which count as encumbrances of decision, then, is an open set which will accommodate any condition shown to have the indicated effect on decision-making.

Decisions supportive of intervention. The Principle of Paternalism allows intervention only when the hypothetical unencumbered decisions of the person to be coerced are *supportive of the intervention*. This means that if the person were not in the encumbered circumstance, he would expressly consent to the intervention. Thus, if there is good evidence that the person, when in an unencumbered state, would agree that if he were in circumstances of the sort in question, he should be, and would want to be, subjected to paternalistic coercion, then that person's decisions may be said to be supportive of intervention.

Hypothetical unencumbered decisions. The most difficult concept in the Principle of Paternalism is that of hypothetical unencumbered decisions. The justification of intervention requires that the intervening party have some information as to what the person's decisions or choices *would be if* they were not encumbered. Sometimes this information is not difficult to obtain. Suppose that White is about to take a drink of poisoned wine not knowing that it is poisoned. White's good friend Black, who knows that the wine is poisoned and that White does not know it, must consider whether to intervene. Suppose that Black has every reason to believe that White has no intention of doing away with himself: White, let us suppose, has expressed no such intention; on the contrary, he has expressed the intention of carrying out a wide variety of plans in the future. Under such circumstances, Black surely

has good evidence that White would choose not to drink the wine if he knew it was poisoned and that White would support intervention to prevent him from doing so in the encumbered circumstance. Thus, according to the Principle of Paternalism, Black's intervention would be justified.

The unique feature of the case of Black and White is the fact that the prediction about White's hypothetical unencumbered decisions can easily be verified. Black intervenes, tells White that the drink is poisoned, and then if White expresses no further desire to drink the wine and no objection to the intervention, Black's prediction is verified. In this case verification is not difficult to obtain because the encumbrance of White's decision—ignorance of the fact that the wine is poisoned—is fairly easily removed.

Other cases are more problematic. Sometimes an intervening party, unlike Black in the above example, will have no personal information about the person to be coerced. In such cases, prediction of the hypothetical unencumbered will is more difficult. In general, the smaller the quantity of information about the person to be coerced that is available to the intervening party, the more difficult it will be to make accurate predictions about the former's decisions in the absence of the encumbrance. Another factor which serves to make accurate predictions difficult is the nature of the encumbrance involved. When a person is ignorant of one relevant fact about what he is doing, as in the case of the poisoned wine, it is relatively easy to predict what effect the elimination of the ignorance will have on the person's decisions. However, when the encumbrance is something like emotional stress or possibly mental illness, it is much more difficult to predict the effect of the removal of the encumbrance. If, for instance, the case involves someone who has been mentally retarded all of his life, it would be impossible to make detailed predictions as to what his view would be with any great accuracy. Thus, both the lack of clear-cut information about the effect of removal of the encumbrance and the lack of personal information about the party to be coerced are factors contributing to inaccuracy in the making of predictions as to the nature of a person's hypothetical unencumbered decisions.

The determination of a person's hypothetical unencumbered decisions requires, however, that all information which may have a bearing on these decisions be taken into account. This includes information about the person's goals, beliefs, and

values, and also information about the probable effects of various forms of encumbrance on a person's decisions. Much of the information required, therefore, is of an empirical nature and is thus subject to the difficulties of making accurate predictions on the basis of empirical evidence. The question of whether a given act of paternalism is justified may thus often be in doubt due to the uncertainty of the prediction about the person's hypothetical unencumbered decisions.

It will be noticed that the Principle of Paternalism permits, in cases in which paternalism is permitted, overriding or ignoring the choice which the person coerced actually expresses at the time of the intervention. Paternalism often involves overriding or ignoring such choices (an exception would be the case of unconscious persons), and so a question arises as to what overrides them. We have seen that Dworkin answers this question by appealing to the choices of "fully rational individuals." Dworkin's approach exemplifies the view which substitutes what I shall call the "rational will" for the will of an individual. This is *not* the position expressed by the Principle of Paternalism. That Principle can therefore be further explicated by comparing its appeal to hypothetical unencumbered decisions to the appeal of other views to the rational will.

IV

Allow me to introduce some additional terminology at this point. The choice, decision, or will that a person actually expresses at any given time I shall call his *empirical* choice, decision, or will. Paternalism thus normally involves overriding or ignoring a person's empirical will. But in favor of what? One possibility here, as we have seen, is the appeal to the rational will. By the "rational will" is meant the will which would be expressed by any fully rational being, as determined in abstraction from the individual characteristics of any such being. In general, the rational will is identical with whatever choice is required by reason as such, and the determination of the rational will requires consideration of the concept of rationality. However, as indicated above, this is not the appeal made by the Principle of Paternalism.¹⁷

Rather than appealing to the rational will, the Principle of Paternalism appeals to what the individual's empirical will *would be* if it were unencumbered. The individual's empirical will is overridden, but in favor of that same individual's empirical will as it would be in more favorable circumstances. In other words, the Principle of Paternalism requires that it be determined what the individual's actually expressed, empirically observable will would be in an unencumbered situation. One significant aspect of this is the fact that any specific claim about the nature of a person's hypothetical unencumbered will is always empirically verifiable in principle. To verify such a claim, all that is required is the determination of the person's empirical choice at a time when his will is unencumbered. In some circumstances, there will be no problem about this. To verify a claim about the unencumbered will of a person whose will is encumbered by ignorance of the fact that his glass of wine is poisoned, one need only inform him about the poison and observe the decisions he expresses in light of that knowledge. In some cases, however, it is not possible to remove an encumbrance at will. For example, we lack the ability to eliminate mental illness at will, and so if that is the encumbrance in question, it may not be possible to actually verify a claim about a person's hypothetical unencumbered will. But such verification remains possible in principle.

In this way the concept of hypothetical unencumbered will differs from that of the rational will. The rational will is not something which can be determined empirically. Empirical facts play a role in the determination of the rational will only as premisses, if at all. That is, determination of the rational will may require knowledge of general facts about human beings, as in the Rawlsian original position, but empirical facts cannot be used to check whether the correct conclusions have been drawn. Moreover, the determination of the rational will is done in abstraction from the individual characteristics of any person, while the determination of a person's hypothetical unencumbered will could not correctly be made without considering such characteristics.

But why should the Principle of Paternalism

¹⁷ It must be noted that the use of the concept of the rational will discussed here is distinct from its use as a method for the justification of moral principles. As used to override the empirical choice of some individual, the rational will is used as the "real" will of that individual. Used as a method for the justification of moral principles, there is no substitution of the rational will for the empirical will of any individual. Thus, while the same conclusions about the nature of the rational will could be used in either of these two ways, it must be noted that they are distinct uses and that only the substitution of the rational will for the empirical will of an individual is being questioned at this point and in this way.

rely on the hypothetical unencumbered will rather than on the rational will? The answer is that this use of the rational will would constitute a departure from the "weak" paternalism of Mill in favor of a "strong" paternalism which Mill would have rejected. Since the present concern is the elaboration of a criterion of justified paternalism which captures the Millian position, the use of the rational will is unacceptable. Let us consider why the rational will would lead to strong paternalism.

Suppose we were to use the rational will rather than the hypothetical unencumbered will as the determinate of a person's "real" will, i.e., the will which is of greatest moral significance. Since a correctly determined hypothetical unencumbered will is an unencumbered empirical will, this would be to give greater weight to the rational will than to a person's unencumbered empirical will. But if the rational will can outweigh a person's unencumbered empirical will, there is no restriction of intervention to cases in which the will of the person coerced is encumbered, for the greater weight of the rational will means that it should be followed. Thus, if the rational will were used, even someone whose decisions were not encumbered in any way could be coerced, so long as the coercion were in the direction indicated by the rational will. There would therefore be no restriction of paternalism to cases in which a person's ability to make decisions is hindered in some way, and therefore no approximation of Mill's position.

Further, the use of the rational will would provide no way to ensure that eccentric behavior would not be the grounds for paternalistic coercion. Since the rational will is determined in abstraction from the characteristics of any individual, it is possible that some forms of paternalism would be acceptable to the rational will but not to the unencumbered empirical will of an atypical person. Since the use of the rational will could thus permit paternalistic coercion of persons with unusual views because their views would not be represented in the determination of the rational will, the use of the rational will fails to protect the minority view which Mill was so concerned to protect.

Despite the fact that the Principle of Paternalism requires giving preference to the unencumbered empirical will over the rational will, there is a role for the rational will in the application of the Principle of Paternalism, even if only a secondary one. The hypothetical unencumbered will rather than the rational will is to be favored in the application of the Principle, but in some circum-

stances it will be impossible to determine the former with any high degree of certainty. In such cases the rational will can be taken as *evidence* in the determination of the direction of the hypothetical unencumbered will. Suppose, for example, that a potential intervening party is dealing with someone who has been in an accident, was knocked unconscious, and is in need of surgery to avoid death. Suppose further that no personal information about the victim can be obtained. Then, assuming that the rational will would favor actions necessary to avoid certain death, an appeal to the rational will could be made to support the claim that the surgery would be allowed by the person's hypothetical unencumbered will. When no evidence but that provided by the rational will is available, the rational will would be decisive; when other evidence is available, the rational will would be only one factor and would not be decisive. Thus, the rational will does potentially have a secondary role in the application of the Principle of Paternalism.

V

I should like to conclude by briefly considering how the Principle of Paternalism would deal with the examples and considerations discussed in section I. We saw there that the categories of Ignorance, Emotional Stress, Compulsion and Undue Influence, Mental Illness, and Non-rationality all seem to represent states of mind under which actions may be performed or attempted. In terms of the Principle of Paternalism, these categories all constitute *encumbrances*. They are all conditions or states which may affect decision-making adversely by causing persons to make decisions they later regret. Because these conditions constitute encumbrances of decision, when one or more of them is present the first condition of justified paternalism is met. The second condition is met only if there is good evidence that the person would support intervention if he were free of all encumbrances. This condition demands that the particular circumstances of each case be considered, but it is clear that there are possible cases in which a person would, perhaps to avoid harm to himself, support intervention.

The Principle of Paternalism also provides a way to account for the importance of considerations about the harm a person is likely to cause himself. Most discussions of paternalism, as we

have seen, attribute great importance to the threat of severe and permanent harm, and the Principle of Paternalism is no different. However, under the Principle of Paternalism, the role played by the threat of serious harm is not that of a necessary or sufficient condition of justified intervention, but of evidence of the satisfaction of such conditions. That is, when a person's actions threaten to bring serious harm to him, this constitutes evidence both that an encumbrance is present and that the person's unencumbered decision would be different.

When a person does something likely to cause himself serious harm, this constitutes some evidence that the person's decisions are encumbered because the presence of an encumbrance is one possible explanation of the fact that the person is about to harm himself. He may be harming himself, for example, because he does not realize what he is doing. However, because of the possibility that a person could make an unencumbered choice to harm himself, this is not conclusive evidence that an encumbrance exists. In fact, the threat of serious harm is likely to constitute sufficient evidence of an encumbrance only in situations in which the immediacy of the threatened harm prohibits further investigation.

The threat of serious harm constitutes evidence that a person's hypothetical unencumbered decisions would be supportive of paternalistic intervention because of the fact that virtually everyone wishes to avoid serious harm virtually all of the time. Consequently, it is highly probable that anyone whose decision is encumbered and who is threatened with serious harm would support actions taken to prevent such harm. However, this evidence is again not conclusive, since personal

information about the person to be coerced could indicate that he does wish to inflict serious harm on himself.

The threatened harm need not, however, be serious in order to provide evidence which may serve to justify paternalistic intervention. Consider again the case of the person about to drink the harmless, foul-tasting liquid which looks like wine. When a potential intervening party has every reason to believe that the person does not know that the liquid is not wine, intervention can be justified because of the high probability that the person involved is like most other persons in that he does not desire to drink foul-tasting liquids. Here the unpleasantness to be avoided is slight, but paternalism can be justified because of the strong evidence as to the nature of the person's hypothetical unencumbered choice. Minor inconveniences can thus constitute grounds for paternalistic intervention in some circumstances.

The Principle of Paternalism thus has two important virtues not shared by other proposed criteria for the determination of when paternalistic interventions are justified. First, its application yields conclusions which are in agreement with common intuitions about what sorts of cases are such that paternalism should be allowed. Second, it can account for the relevance of the various factors clearly significant for the justification of paternalistic interventions in a way which is highly plausible in that it provides for the organization of these factors without ignoring the differences between them. Thus, for capturing intuitions of the sort shared by Mill, the Principle of Paternalism is the best criterion for the justification of paternalism.

VII. THE JUSTIFICATION OF MORALITY

DAN W. BROCK

IN just what senses morality, or acting for moral reasons, requires or admits of rational justification is an issue that has long divided moral philosophers. In this paper I will argue that it is not possible to provide practical reasons that justify morality or acting for moral reasons. And while it is tempting then to conclude that there is not good reason to act morally, that would beg the second issue I want to discuss here—in what sense morality requires further justification. I shall argue that it is a mistake to assume that it requires justification. Finally, I shall offer some very brief remarks about the nature of a morality in order to clarify how what can be said in support of being moral should be understood.

I

I need to provide a brief account at the outset of what I understand by morality, or a morality, so that the nature of the issue of its justification will in turn be clear. At different times I shall speak of morality, or a person's morality, as a system of rules (or principles), judgments or reasons that are practical or action-guiding; that is, they state that actions of a specified sort ought, or ought not, to be performed. This first condition, that moralities are practical or action-guiding, simply locates morality within the domain of practical reasoning.

The second feature of a morality concerns the content of its action-guiding rules, judgments or reasons. Since the issue of justification I am interested in here can be raised in essentially the same form for any prominent moral view or normative ethic, this feature must be common to, for example, utilitarianism and theories that oppose utilitarianism from considerations of justice and fairness. The contrast I want to draw in characterizing morality is not between different normative ethics, but between most any moral

view and explicitly non-moral systems of practical reasoning. I shall do this by specifying only that the content of a morality is such that it requires a person to act in ways that may conflict with the maximization of his own self-interest; in any morality, some consideration other than the agent's own self-interest justifies action, and can in at least some possible cases of conflict, override and so require the sacrifice of the agent's self-interest. I want to focus on this requirement, shared by different moral theories, of some possible sacrifice of one's own self-interest since it is this requirement that is usually thought to place morality in need of justification.¹

Characterizing moralities in this way brings out two points. First, it suggests that the principal non-moral alternative to morality among systems of practical reasoning will be one or another version of egoism. When I refer with no further qualification either to a non-moral system of practical reason or to egoism, I shall have in mind universal egoism: For any person *P* and action *a*, *P* ought to perform *a* if, and only if, doing so will maximally promote his overall self-interest. Second, it brings out the final feature I shall use in characterizing morality, often called the finality, overridingness or authority of morality: moral judgments (reasons, rules) override non-moral judgments in cases of conflict, and specifically, override judgments of self-interest.²

It is essential for fully appreciating the problem of justifying morality that morality's characterization include both a condition on its content and on its finality. If morality is characterized as final but its content left open, it is difficult to raise the question of its justification at all. If morality is specified simply as a system of judgments of what, all things considered, one ought to do, then the question whether one ought to be moral—do what,

¹ For an analysis of the concept of morality, this content condition that moral requirements sometimes conflict with requirements of self-interest would have to be specified differently; in this form, it could be satisfied by a system of practical reasoning that we would intuitively regard as non-moral. However, specifying it in this way helps focus the justification issue I am concerned with without, so far as I can see, affecting my argument concerning morality's justification.

² I state the finality or overridingness condition of morality as a property of moral judgments, reasons, or rules because moral philosophers making judgments, reasons, or rules primary in their theories have all adopted it, though I believe it is problematic whether finality can be successfully defended as a property of moral reasons.

all things considered, one ought to do—makes no sense at the outset. With my characterization of morality as well, the finality condition raises crucial difficulties for justifying morality. But at least if we also give a minimal specification of the content of moral reasons (rules, judgments) we can then sensibly ask why *these* particular reasons, reasons with *this* content, should be final or overriding. And it is, of course, only when morality is understood as having at least minimal content of the sort specified above that the issue of the justification for acting morally or accepting moral reasons is raised.

II

How might morality, or acting morally, be given justification? I shall use the distinction between providing a moral and providing a non-moral justification of morality, and in this section take up the former possibility. While I have nothing new to add on it to the points made by Singer, Toulmin and others, it will be helpful for what comes later to have the difficulty clearly, though briefly, before us. Consider a person contemplating whether to murder his rich parents in order to inherit their money. Suppose he concludes that killing another person is *prima facie* morally wrong, that in the present case there are no competing considerations that override the *prima facie* wrongness of killing persons, and so finally that this proposed killing is, all things considered, morally wrong. His reasoning to this point remains within the framework of moral reasoning, for not until this point have all relevant moral reasons or considerations been exhausted.

But now suppose he asks whether the fact that his proposed action is morally wrong justifies his not doing it: that is, whether he ought to or is justified in acting on moral reasons. This is the situation of the moral skeptic who wants a justification of morality, of acting morally. As Singer has rightly put it, there is no further moral reason he wants or that we could give him—all relevant moral considerations have already been taken account of in his judgment that the proposed action is, all things considered, morally wrong. And—this is the crucial point—no moral reason could justify morality or accepting moral reasons. Anything that could be offered as such, as a moral reason for being moral or accepting moral reasons, would itself be an instance of what he was asking for a justification of, for a reason for accepting.³

The picture of practical reasoning presupposed here is of different kinds of reasons; or different modes or systems of practical reasoning (for example, legal, theological, prudential, as well as moral). A reason or justification for accepting the entire set or system of moral reasons must be external to that system, and so come from some other area or mode of practical reasoning. “Why should I be moral?”—understood as asking for a *moral* reason of justification for being moral—is a meaningless question.

III

The above should help clarify what is meant by a non-moral reason or justification for morality—it must be a reason or justification drawn from some non-moral system (or area, or mode) of practical reasoning. As already indicated, the non-moral, external standpoint from which I shall explore the possibility of a justification of morality is that of a person's own rational self-interest. However, I shall also show how the argument against the possibility of justifying morality in terms of rational self-interest can be generalized to cover any other attempted non-moral justification of morality as well.

What is it about morality as part of a system of practical reasoning that makes its justification impossible? Is it something special about morality, or is it likewise impossible to justify various non-moral elements of practical reasoning as well? An argument exactly analogous to that of the last section would show that a legal justification of legal reasoning, a prudential justification for accepting prudential reasons, and so forth, is impossible. But, on the other hand, it is not at all clear why an external justification for accepting legal, prudential, etc. reasons is not possible, whether or not it is in fact forthcoming. For example, suppose we consider the law in a particular country at a particular time as having a more or less well specified content, a content that provides a system (or part of a system) of reasons for acting. It is then perfectly sensible to ask whether accepting and so acting on this particular system of legal reasons is justified by or within another framework of practical reasoning such as self-interest or morality. (Whether, and for what reason, this question is important is, of course, a separate issue.) Assuming the system of legal reasoning in question is logically distinct from

³ Marcus Singer, *Generalization in Ethics* (London, 1963), p. 319.

that of, say, self-interest, its justification will require appeal to various contingent facts whose truth will guarantee that following this particular system of legal reasoning is in one's self-interest. This is a natural way of making sense of the demand for an external justification of a system of practical reasoning, and one which seems applicable to moral systems as well as non-moral systems; I shall examine this possibility shortly. It confirms what is commonly assumed—that external justification of a specified system of legal reasoning, of economic reasoning, and so forth, is not *in principle* impossible. And it as well suggests that if a non-moral justification of morality is in principle impossible, this is not due to the content of the reasons in a system of moral reasoning, as opposed to the content or nature of legal, economic, etc. considerations.

Rather, it is the formal quality of moral judgments—that they are final or overriding—that creates the difficulty for external or non-moral justification of morality, and this is a property that is not, and cannot be, shared with non-moral practical judgments. The property of moral judgments I call finality or overridingness is not always explicitly identified in this way, but it is widely accepted. For example, it is a property of the principles that define a conception of right in Rawls' theory of justice and is a property of moral judgments entailed by their prescriptivity on the otherwise rather different theory of Hare. Now assume that the property of finality is a necessary condition for a principle to be a moral principle. Could a non-moral justification of accepting and in turn acting on moral principles then be provided? To say that moral principles are final or overriding is to say that they are the final and decisive appeal in practical reasoning, that they specify what, all things considered, we ought to do. Their finality specifies *with finality* what we ought to do, not what we ought to do only *if* it is also legal, in our self-interest, etc. But if morality requires justification in terms of self-interest, establishing that under specified conditions acting morally is in our self-interest could at most support the position that *when* acting morally is in our self-interest, then acting morally is justified. Thus, any such justification must necessarily fail to justify an essential feature of moral judgments—their finality or overridingness.

It might still be thought that an external or non-moral justification of morality, consistent with maintaining its finality, is possible along the lines suggested above for an external justification of a

system of legal reasoning. By appealing to certain contingent truths, we could try to show that, when applied in the world as it is, some non-moral system of practical reasoning, in terms of which morality is to be justified, yields precisely the same results as the moral system; applied in the world as it is, the two systems are extensionally equivalent. (In fact, a somewhat weaker claim—that all the particular judgments yielded by the system of moral reasoning were included within, but did not exhaust, the set of judgments yielded by the system of non-moral reasoning—would be adequate; the argument which follows holds for this possibility as well.) This would allow justification of the moral system by appeal to the non-moral system, and is consistent with denying that moral judgments are ever overridden by non-moral judgments. Moral judgments are not overridden by judgments of prudence or self-interest, the relevant non-moral system, only because certain contingent truths about the nature of the world guarantee that the two kinds of judgments never in fact conflict.

It remains, nevertheless, logically possible for there to be conflicts between moral and non-moral judgments, specifically judgments of self-interest, whether or not any such conflict in fact arises in the world as it is. But the property of finality or overridingness possessed by moral judgments should be understood as entailing that in any such logically possible case of conflict between a moral judgment and a non-moral judgment, for example of self-interest, the moral judgment is overriding. Being a moral person, adopting the moral point of view, involves being prepared to do what morality requires not just when its requirements happen to coincide with one's self-interest, but at all times. If the finality or overridingness condition means only that moral judgments are not overridden by any non-moral judgments, such as those of self-interest, because the world is happily so-arranged that morality and self-interest never in fact conflict, it would be pointless ever to insist on the finality condition. The condition of overridingness would apply to moral judgments only when and if the world were such that it was unnecessary and never came into play in practical reasoning. It is perhaps worth adding that the world seems to be such that moral requirements do sometimes conflict with one's overall self-interest, but the point is that even if they did not, the happy coincidence of morality and self-interest could not fully justify the finality of moral judgments.

I have been assuming that the property of

finality or overridingness is a necessary, though not a sufficient, condition of a system of practical reasoning (and the principles, judgments and reasons which comprise and follow from it) being a moral system. Some philosophers, however, have rejected finality as a necessary (or sufficient) condition of a morality. They generally characterize moral judgments as judgments having a particular logical form, e.g. universalizable or formulated in universal terms, and/or a particular content, e.g., making reference in some way to the satisfaction of human needs, interests, etc. It is then thought to be logically open whether morality ought to be taken as overriding in practical reasoning, and this seems a way of making sense of the demand for a justification of morality not precluded by any arguments of the sort I have offered above. But it is by no means clear just what sense is to be given to the request for a justification for taking morality as final. It would, of course, be possible to show what kind of judgments, for example moral or self-interested, are in fact taken as final in some particular person's system of practical reasoning. But how would the fact that he took, say, moral judgments as final *justify* his doing so, or show that he *ought* to do so? Suppose that he has not yet decided which kind of judgment to take as final, and asks which he ought to take as such. Any answer will have the general form: whichever kind of judgment or practical reasons, moral or self-interested, possesses property *p* ought to be taken as final or overriding. But this seems to make *p* the ultimate or final consideration justifying action, not the moral judgment, since it is only by possessing *p* that the moral judgment justifiably overrides the prudential or self-interested judgment. We could only show someone that moral judgments, say, ought to override self-interested judgments, even though he in fact took self-interested ones as overriding, by showing that the property *p* that he believed self-interested judgments possessed and that justified their finality was in fact possessed by moral and not self-interested judgments. But this would still leave the possession of *p*, and not moral considerations themselves, the ultimate or overriding considerations in his system of practical reasoning.

Since many philosophers seem to have believed that if finality is not assumed to be a necessary property of morality, it is then possible to show that morality ought to be accepted as final, it is instruc-

tive to look at just how one such attempt fails in Kurt Baier's book *The Moral Point of View*. Baier seeks to demonstrate that moral reasons are "really superior to reasons of self-interest," that there is "a reason for regarding moral reasons as superior to all others."⁴ His claim is that "our very purpose in 'playing the reasoning game' is to maximize satisfactions and minimize frustrations"; this purpose in turn provides the criterion—best serving the interests of everyone—by which reasons are to be assessed as good reasons or not, and by which the hierarchy of reasons is to be established and justified.⁵ He then argues that everyone's interest is better served in a world where everyone accepts moral reasons as overriding reasons of self-interest than in a world where everyone simply follows reasons of self-interest. Thus, moral reasons really are superior to and ought to override reasons of self-interest.

But consider more carefully what Baier takes to be the final or overriding consideration in practical reasoning. This must be not moral reasons, but something like "best serving persons' (everyone's) interests." It is at least possible that everyone's interests would be better served if everyone followed only reasons of self-interest rather than morality. If so, Baier is committed to the position that moral reasons should not override self-interest. His claim is only, and can at most be, that it is contingently true that everyone's interest is in fact best served by everyone's accepting moral reasons as overriding self-interest. The finality or overridingness of morality, however, implies that one ought to act morally regardless of any non-moral considerations. Baier's argument, on the other hand, makes the finality of morality contingent on just such a non-moral consideration—its best serving everyone's interests. Thus, he fails to establish morality's finality.

Baier would probably reject my interpretation of his position and maintain instead that on his view moral reasons are final or overriding, while best serving everyone's interests is only the purpose of practical reasoning and so the criterion for putative reasons being good reasons. Showing that accepting moral reasons as final best serves everyone's interests, Baier could then claim, provides the justification of morality's finality we have been looking for—it is justified by its promotion of the purpose of practical reasoning. Such a response goes

⁴ *The Moral Point of View* (New York, 1965), pp. 148–149.

⁵ *Ibid.*, p. 141.

wrong in its implicit assumption that, by appealing to the supposed purpose of practical reasoning, we somehow step outside practical reasoning itself, and so can then use that purpose to justify our system of practical reasons having the content and hierarchy it does. Even should appeal to the purpose of practical reasoning enable decisive determination of the content and hierarchy of the system of practical reasons, it is enough to point out that there are alternative plausible conceptions of the purpose of practical reasoning that would determine or justify systems of practical reasons with a different content and hierarchy: for example, Rawls' notion of providing a fair adjudication between the competing claims persons make on each other and on their common institutions, or the egoists' notion of *each* person's maximizing satisfaction of his own interest. Proposing a specific purpose of practical reasoning amounts to proposing a general and ultimate practical principle, from which the more specific content of a system of practical reasoning is to be derived. But any moral skeptic who questions why he should accept this system of moral reasons as final in practical reasoning, can and no doubt will question the acceptance of that particular conception of the purpose of practical reasoning as well. In this sense, the purpose of practical reasoning is just as much in need of justification, and just as much a part of that for which justification is being sought, as is the system of moral reasons derived from it. Appeal to purposes cannot settle the issue raised by the moral skeptic who asks for a justification for accepting a system of moral reasons as final in practical reasoning.

If I am right that it is not possible to justify accepting morality, whether its finality in practical reasoning be taken as a necessary or contingent feature of it, this suggests that it is only possible to *describe* the structure of a person's system of practical reasoning, and to determine whether, in it, moral reasons (principles, judgments) are in fact final or overriding. There is no way to show as well that morality ought to be accepted as final in practical reasoning: if it is accepted as final, there will be no superior considerations to appeal to to justify its finality; if it is not accepted as final, no appeal to final, non-moral considerations could justify morality's finality. I shall explore briefly the import of this conclusion in Section IV.

There is an additional difficulty in supposing that accepting morality requires, and could receive, justification in terms of a person's own self-

interest, or what I shall call for short, expediency. It is common to distinguish acting morally from acting expediently. For example, a person may not cheat another simply because it would be cheating and he believes cheating is wrong—there is a duty not to cheat; or he may not do so because he believes there is too great a risk of being caught—it is in his interest or expedient not to cheat. Now in a sense, in each case, he performs the same kind of action—he does not cheat—but many philosophers would insist that only in the former case does his action have moral worth, does he exhibit a good will and deserve moral praise. But suppose in order for acting morally to be justified, for moral reasons really to be good reasons for acting, it must be expedient to act morally. This means that the only good moral reasons are those it is expedient to act on as well, and that moral requirements can never conflict with (and presumably are a proper subset of) the requirements of expediency. Since in every case in which good moral reasons and reasons of expediency apply, they require the same action, how can we now distinguish morality from expediency?

The natural response is that the motives of the moral person and the expedient person remain different: the moral person performs the action because or for the reason that it is morally right, while the expedient person performs it because or for the reason that it is expedient or serves his self-interest. But if morality requires and receives justification in terms of expediency, we cannot distinguish the two in this way, for the moral person has become the expedient person. The person who acts on a moral reason only accepts that reason as a good reason or justified basis for action because it has been shown that morality is expedient. His system of moral reasons, to be worthy of motivating, must conform to the requirements of expediency. We can no longer distinguish one person as doing the act just because it is morally right since, though he may have the moral reason "before his mind" when acting, that reason only is a good practical reason for him because acting on it is expedient. The property of the action that he believes ultimately justifies his performing it, is its expediency, not its morality. His position is analogous to that of the act utilitarian who adopts rule-of-thumb decision-making guides such as "Don't cheat," because doing so will be more likely actually to produce the maximization of utility in his actions than will applying the utilitarian standard case by case. However effective such rules-of-thumb may

be as decision-making aids, their adoption does not result in a standard of rightness, or system of practical reasons justifying action, different from act-utilitarianism. Precisely the same is true if morality is justified only if, and to the extent that, its adoption is expedient or maximally promotes one's self-interest—the adoption of morality does not produce a standard of rightness justifying action different from expediency or the promotion of one's own interest. And if the person who acts on these "moral reasons" has a standard justifying action no different from the expedient person, then there is no basis for singling out the former but not the latter as a good person, possessing moral worth or a good will, and deserving of moral praise. Thus, if we are to maintain the distinction between morality and expediency, we cannot require justification of morality in terms of expediency or self-interest.

IV

If it is not possible to appeal to a moral or non-moral reason to justify morality, or accepting moral reasons at all, is accepting and acting on moral reasons then irrational and unjustified? The question is whether a justification, and specifically a non-moral justification in terms of the agent's own interest, is needed, rather than possible, for morality. Since Plato, many philosophers have assumed that, without such justification, a person does not have good reason to be moral, and so have usually tried to show that acting morally best promotes a person's own interest or good. This assumption, however, is question-begging in a crucial respect. To assume that morality requires non-moral justification if we are to have good reason to act morally, is to assume that practical reasons are ultimately all non-moral, and so moral reasons must be reducible to non-moral reasons if they are really reasons at all. More specifically, to assume that morality needs to be shown to be in one's interest is to assume that practical reasons are ultimately self-interest. An alternative possibility is a broader system of practical reasons that specifically includes more reasons than promoting one's own interests.

To make this more precise it will be helpful to distinguish what is a reason for an action in a strict sense, which I shall call a basic reason, from what I will call derivative reasons. A basic reason in some person's system of practical reasoning is a property of an action that, under the given

description, in and of itself is a reason for or against performing the action. A derivative reason in that person's system of practical reasoning is a property of an action that is only a reason for or against performing that action when some further premiss, information, etc. is added. To assume that morality requires non-moral justification and so that practical reasons are ultimately non-moral, or specifically promoting one's own interest, is to assume that the only basic reason is promoting one's own interest. In such a view, promoting Jones' interest can be a basic reason only for Jones; it can be a derivative reason for Smith, but only if the additional premiss that promoting Jones' interest in turn promotes Smith's interest is true. The alternative I am suggesting is that one's own interests are not the only basic reasons in one's system of practical reasoning. The clearest contrast with the egoist's view that the only basic reason is promoting one's own interest is the utilitarian's or consequentialist's view that promoting some person's (anyone's) interest is a basic reason for action for any particular person (everyone); that an action promotes another's interest gives me a reason for doing it directly, whether or not promoting that other person's interest in turn promotes my own interest. A deontologist holds that an action being of a certain type, such as the keeping of a promise or the killing of a person, gives anyone a basic reason for performing or not performing it. So if moral reasons as well as self-interested reasons are basic reasons, then it is a mistake to hold that acting morally must be justified in terms of self-interest, or any other non-moral consideration, for there to be good reason to act morally.

Notice that in regard to one's own wants or interests, it is commonly assumed that if a person wants *x*, or if *x* serves his interest, this *in itself* gives him a *prima facie* reason to promote *x*. However, it is possible to demand of our own wants or interests something analogous to what the moral skeptic demands of morality—that promoting our own wants or interests must be justified by some other non-self-regarding reason or consideration in order to provide us with a good reason for action. For most persons, the proper response to this demand is that promoting our own wants or interests requires no further justification to be a good reason—it is a basic reason. If Jones wants to eat, if eating is in his interest, *that* gives him a reason to eat; nothing more is needed for him to have a good reason to eat. Nor is it supposed the promotion of our own wants or interest is in any

way undermined as a reason for acting because no further justification for it in terms of something else is forthcoming. I am simply suggesting that moral reasons can have the same "ground floor" status, as basic reasons, in one's system of practical reasoning; to be good reasons, they require no more justification in terms of something else than promoting our own wants or interests does.⁶

This should make clearer why I claimed earlier that it is only possible to describe, but not to justify, any particular system of practical reasoning. We can describe which reasons are basic reasons, and the hierarchy of reasons including which reasons are final or overriding in that system, but no further practical reasons exist to which we could appeal to justify accepting either basic reasons as basic or overriding reasons as overriding. The impossibility of justifying morality then results from the place of moral reasons as both basic and overriding (different properties of reasons) in a system of practical reasoning and from the familiar point that a process of justification must reach or stop with some justifying reason that is not itself justified by any further reason. In the face of this it is tempting to claim that not only which basic moral principles a person accepts, but whether he accepts any moral principles at all, is a matter of personal choice, decision or commitment. One must just choose morality, or love justice or morality as it is sometimes put, because no justification for being moral is possible. Nevertheless, without further elaboration, such claims are misleading and do very little to put the moral skeptic's doubts to rest. For if it is only possible to describe a system of practical reasoning in which moral reasons are basic and final, he will rightly point out that he can likewise describe an egoistic or other system of practical reasoning in which moral reasons are absent. And he will then want to know if there is any sense in which morality is worthy of our commitment or love. Given that no practical reason can fully justify being moral, what is the nature of morality's appeal for us?

A full answer to this question cannot be attempted in this paper. However, a few brief remarks will indicate the direction I believe such

an answer should take. Moral philosophers have commonly attempted to meet the challenge of egoism by arguing that egoism in its different forms is internally inconsistent, or that it fails to be a *rational* system of practical reasoning, or that it fails to be a *theory* of practical reasoning, or that it does not meet the requirements for being a possible *moral* theory. It is common as well to try to show that the adoption of a moral system of practical reasoning is a rational *necessity*. No such attempts have been, in my view, fully successful. They have failed to meet many of the substantive issues posed by the moral skeptic, and have advanced an inadequate conception of the nature and grounding of theories of practical reasoning. What is needed, I believe, is a coherence account of theories of practical reasoning and their justification. This will involve developing the full nature and implications of each theory of practical reasoning by working out the logical and empirical connections between its various parts. In doing so we will only be providing a full description of each theory, not showing it to be a rational, or logical, or theoretical, or moral necessity.

A common objection to a coherence view of systems of practical reasoning is that it seems to commit one to some form of relativism, and so provides morality with too weak a foundation, too weak to hold on us; it fails to account for the inescapability of moral requirements. But this is not, I think, a result of the coherence view itself. It comes primarily from too narrow a conception of what such a coherence account should include. Besides the usual specification of the general moral principles, and the particular judgments they imply, a coherence account of a theory of practical reasoning should include at least the following: a framework of justification (as in the ideal observer view, or Rawls's original position);⁷ a theory of moral psychology, that is, an empirical account of how persons will come to desire to act in accordance with the normative principles of the theory;⁸ the identification of particular emotions or feelings, such as guilt and shame, logically or empirically related to acceptance of the system of practical reasoning in question;⁹ the identification of par-

⁶ My point here does not concern the relation between a person's wants and interests, and I should not be understood as suggesting that the relation is one of simple equivalence.

⁷ See, for example, Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research*, vol. 12 (1952), pp. 317-345 and John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971), ch. 1.

⁸ A useful review of the literature of moral psychology is Derek Wright, *The Psychology of Moral Behavior* (Middlesex, 1971). Rawls, *A Theory of Justice*, *op. cit.*, ch. 8, draws heavily on the work of Jean Piaget and Lawrence Kohlberg in developing a theory of moral psychology for his principles of justice.

⁹ See, for example, *Guilt and Shame*, ed. by Herbert Morris (Belmont, 1971).

ticular social practices, such as praising and blaming, giving or seeking moral advice, etc., consistent with or precluded by the theory of practical reasoning in question;¹⁰ the identification of non-moral social relations such as friendship and trust dependent upon or incompatible with the theory;¹¹ and finally, the identification of certain beliefs, e.g. the metaphysical belief that we are one person among many equally real persons in a single world, that commit us to practical principles of a certain sort.¹²

All of this is, of course, sketchy; I hope to work it out at greater length elsewhere. I have suggested this wider than usual view of the nature and implications of a system of practical reasoning only to show why I believe the coherence view of theories

of practical reasoning provides such theories, and morality in particular, with a stronger grounding than is commonly thought. For though we can only describe with as much precision and rigor as possible all that is involved in, for example, adopting a morality or adopting some form of egoism, the more that is at stake in doing so, in being a moral person or an egoist, the clearer it may become that we have little if any choice in whether to adopt a morality. If this quite broad view of all that is involved in adopting some particular system of practical reasoning is correct, it is seriously misleading to speak of a choice or commitment to adopt such a wide-ranging and fundamental part of our nature.¹³

Brown University

Received February 5, 1976

¹⁰ These issues are discussed in many places, including Jesse Kalin, "In Defense of Egoism," in *Morality and Rational Self-Interest*, ed. by David Gauthier (Englewood Cliffs, 1970).

¹¹ Helpful here is Charles Fried, *An Anatomy of Values* (Cambridge, 1970).

¹² I have in mind here arguments like Thomas Nagel's in *The Possibility of Altruism* (Oxford, 1970).

¹³ I benefited from helpful discussions of an earlier version of this paper read at Syracuse University and at Dartmouth College.

VIII. A COGNITIVIST SOLUTION TO NEWCOMB'S PROBLEM

RAYMOND DACEY, RICHARD E. SIMMONS, DAVID J. CURRY,
JOHN W. KENNELLY

I. INTRODUCTION

NEWCOMB's problem, a purportedly paradoxical decision problem, has provoked a number of proposed solutions in diverse journals since first published by Nozick [1969]. This problem, which has a superhuman (if not supernatural) "being" as one party, is notable for the fact that none of the analysts (except a well-known novelist) are satisfied with their own solutions. Suspicion of paradox persists (evidence that the problem is not solved). We offer a review of the discussion and yet another proposed solution. Our satisfaction is somewhere between that of the novelist and the others.

Nozick [1969] presents Newcomb's decision problem as follows:

There are two boxes, (B₁) and (B₂). (B₁) contains \$1,000. (B₂) contains either \$1,000,000 (\$M), or nothing. What the content of (B₂) depends upon will be described in a moment.

(B₁) (\$1,000) (B₂) $\left\{ \begin{array}{l} \$M \\ \text{or} \\ \$0 \end{array} \right\}$

You have a choice between two actions:

- (1) taking what is in both boxes
- (2) taking only what is in the second box

Furthermore, and you know this, the being knows you know this, and so on:

- (I) If the being predicts you will take what is in both boxes, he does not put the \$M in the second box.
- (II) If the being predicts you will take only what is in the second box, he does put the \$M in the second box.

(If the being predicts that you will randomize your choice, . . . then he does not put the \$M in the second box.)

The situation is as follows. First the being makes its prediction. Then it puts the \$M in the second box, or

does not, depending upon what it has predicted. Then you make your choice. What do you do?

(Nozick [1969], pp. 114-115, p. 143)

Nozick [1969] and Bar-Hillel and Margalit [1972] both attack the problem with game theoretic techniques, and achieve opposite solutions. Nozick decides that unless you believe in absolute determinism your best response is to take both boxes. Bar-Hillel and Margalit find that selecting the second box only is your best response, regardless of your views on determinism. Others have offered numerous alternative solutions. (See Gardner [1973] and Nozick [1974], especially the latter.)

We note that most of the differences in the several articles on the problem hinge on the interpretation of the probability-words used in the problem statement. Furthermore there are different problem statements, distinguished (in part) by the probability-words used in each statement.¹ Our solution to this problem is based on an analysis of several possible and consistent interpretations of the probability-words. We provide an answer to Bar-Hillel and Margalit's question of ". . . whether the paradox does not lie in the very assumption of such a Being itself." Our short answer is, "yes, it does!"

The first step is to dispose of Nozick's solution. Our argument on this point is equivalent to that of Bar-Hillel and Margalit, but (we believe) more cogent.² First, is it reasonable to entertain such a being, and, if so, in what form? We presume that the being's behavior must be lawlike (in the psychological sense), for if it were chaotic we could never conceive of it. If the being is lawful in this sense, then Nozick's fault is (anthropomorphically) imposing inconsistent laws on the being. Nozick's main argument hinges on the being's making an *irrevocable* move after which we inferior humans can *revoke* our move. Now if this

¹ Levi [1975] expands on this point in his examination of Nozick's analysis of the problem. (Rumor has it that Newcomb gives a problem statement different from those cited.)

² Our argument was developed independently of, albeit later than, Bar-Hillel and Margalit's.

being is to be credibly superior to us (otherwise there is no paradox), we must either be similarly constrained or the being must be able to revoke its moves at least as fast as we revoke ours.³ In either case, Nozick's argument evaporates.

This leaves us with Bar-Hillel's and Margalit's solution, which we will qualify. The second question is whether the being is omniscient (in the sense of determinism) or merely superior. In the former case our decision (the same as Bar-Hillel's and Margalit's) is to say, "*kismet*," and take only B₂. The latter case offers a hope of outwitting the being, but requires precise interpretation of the probability-words in the problem statement. Bar-Hillel and Margalit do not consider this case which our analysis shows to be quite interesting.

We propose to attack the problem from a new side. Where Nozick and Bar-Hillel and Margalit employ game theoretic tools we use elements of contemporary inductive logic. More specifically, we adopt the view of the being as decision maker, analyze some of the cognitivist rules of acceptance/prediction available to him, and develop the conditions of rational behavior for both the being and the human decision maker.

We have noted that most of the expressed interest in the issue of determinism in Newcomb's problem supports the inference that the analysts have a vested interest in demonstrating their own free will. This interest is best typified by Isaac Asimov, the novelist referred to above:

I would, without hesitation, take both boxes. . . . I am myself a determinist but it is perfectly clear to me that any human being worthy of being considered a human being (including most certainly myself) would prefer free will, if such a thing could exist. . . . Now, then, suppose you take both boxes and it turns out (as it almost certainly will) that God has foreseen this and placed nothing in the second box. You will then, at least, have expressed your willingness to gamble on his nonomniscience and on your own free will and will have willingly given up a million dollars for the sake of that willingness—itsself a snap of the finger in the face of the Almighty and a vote, however futile, for free will. . . . And, of course, if God has muffed and left a million dollars in the box, then not only will you have gained that million but *far more important* you will have demonstrated God's nonomniscience. If you

take only the second box, however, you get your damned million and not only are you a slave but also you have demonstrated your willingness to be a slave for that million and you are not someone I recognize as human.

(Asimov, quoted in Nozick [1974], p. 104)

Asimov's view suggests that the appropriate utility structure for the analysis of the problem is one having components for both wealth and free will. In our concluding section we comment on such an "Asimov-utility structure." Inasmuch as several of our cases reveal strategies for outwitting the being, we have dubbed these "Asimov strategies."

II. THE BEING AS DECISION MAKER

So let us agree that the predictor [i.e., being] works as follows: He observes you sometime before you are faced with the choice, examines you with complicated apparatus, etc., and then uses his theory to predict on the basis of this state you were in, what choice you would make later when faced with the choice.⁴

(Nozick [1969], p. 134)

We have a being who observes your present state and, based upon the observation and a theory, predicts your future act. For ease of exposition, let us suppose that the "complicated apparatus" yields the possible observations '+' or '-', the former indicative of selectors of the second box, the latter of selectors of both boxes. On the basis of the observation and theory the being must select one of the hypotheses "The human will take both boxes" (denoted " H_1 " hereafter) or "The human will take only the second box" (" H_2 "). Having made a prediction the being places or does not place the \$M in the second box. What does the being's theory look like, how does it work?

Recall that you have two available actions: you take both boxes or you take only the second box. We view a theory for the being as a mapping from the Cartesian product of {+, -} and the real numbers to { H_1 , H_2 }, determined by the being's (probabilistic) trust in the "complicated apparatus," its initial probabilities of your alternative actions, and a cognitivist acceptance rule. Thus, we view the being in its prediction problem as a classical Bayesian decision-maker who behaves in

³ Our general ability to change our mind in this kind of situation depends on two factors: (1) our ability to perceive the being's changes, and (2) the speed with which we can make necessary computations. The latter factor has been recognized by I. J. Good [1971]. A general treatment of both factors appears possible through appropriate specification of the payoff matrix for an analysis of its *channel capacity* (Shannon and Weaver [1969]) or presence of *requisite variety* (Ashby [1963]).

⁴ This view of the being avoids the operation of backwards causality. If backwards causality is operative, then there is no problem: the human decision maker simply takes the second box only. On both points, see Nozick ([1969], p. 134).

accordance with maximized expected epistemic utility. Later we view it as a more traditional Neyman-Pearson or Bayesian statistician.

III. COGNITIVIST ACCEPTANCE RULES

We consider three rules of acceptance that have received extensive coverage in the philosophy literature.⁵ The first and the simplest is a version of "AGn" (Hilpinen [1968], p. 64) and is called the high probability rule. The rule is as follows:

- (R1) Given a partition $\{h_1, h_2\}$ and evidence e , accept h_1 on the basis of e iff $P(h_1 | e) > P(h_2 | e)$. (AGn)

The remaining two rules are based upon the concepts of information content and relevance. The rules are based upon "Acont" (Hilpinen [1968], p. 108) and a variant of "Acont" (here "Acont'") employing likelihoods (Hilpinen [1968], pp. 112-115, Lehrer [1974]), respectively. The rules are as follows:

- (R2) Given a partition $\{h_1, h_2\}$ and evidence e , select h_1 on the basis of e iff
- $$P(h_1 | e) - P(h_1) > P(h_2 | e) - P(h_2). \quad (\text{Acont})$$

- (R3) Given a partition $\{h_1, h_2\}$ and evidence e , select h_1 on the basis of e iff

$$\frac{P(h_1 | e)}{P(h_1)} > \frac{P(h_2 | e)}{P(h_2)} \quad (\text{Acont}')$$

We are interested in these rules when $\{h_1, h_2\}$ is $\{H_1, H_2\}$ and $e \in \{+, -\}$. The rules yield the following predictions:

- (R1) accept H_1 on $-$ iff

$$\frac{P(H_2)}{P(H_1)} < k-, \quad \text{where } k- = \frac{P(- | H_1)}{P(- | H_2)}$$

accept H_2 on $+$ iff

$$\frac{P(H_2)}{P(H_1)} > k+, \quad \text{where } k+ = \frac{P(+ | H_1)}{P(+ | H_2)}$$

- (R2) accept H_1 on $-$ iff

$$\frac{P(H_2)}{P(H_1)} < K-,$$

$$\text{where } K- = \frac{P(- | H_1) - P(-)}{P(- | H_2) - P(-)}$$

accept H_2 on $+$ iff

$$\frac{P(H_2)}{P(H_1)} > K+,$$

$$\text{where } K+ = \frac{P(+ | H_1) - P(+)}{P(+ | H_2) - P(+)}$$

- (R3) accept H_1 on $-$ iff

$$P(- | H_1) > P(- | H_2)$$

accept H_2 on $+$ iff

$$P(+ | H_2) > P(+ | H_1)$$

Rule (R1) yields unusual results. It is easy to show that $k- > k+$. Indeed, if we presume that $P(- | H_2) > 0$ and $P(+ | H_1) > 0$ (i.e., that the "complicated apparatus" is itself not a "perfect predictor" in the decision theoretic sense), then $\infty > k- > k+ > 0$.

The being operates on the negatively-sloped line $P(H_1) + P(H_2) = 1$. At any point on the line segment

$$S_I = \{\langle P(H_1), P(H_2) \rangle \mid P(H_1) + P(H_2) = 1$$

$$\text{and } 0 \leq \frac{P(H_2)}{P(H_1)} < k+\}$$

the being always accepts H_1 , regardless of the signal. Similarly, at any point on

$$S_{II} = \{\langle P(H_1), P(H_2) \rangle \mid P(H_1) - P(H_2) = 1$$

$$\text{and } k- < \frac{P(H_2)}{P(H_1)} < \infty\}$$

the being always accepts H_2 . Thus, if the being's initial probabilities place it on S_I , then there will be \$0 in the second box; if on S_{II} , then \$M.

A being who follows rule (R1) is open to being beaten by the kind of human, characterized above by Asimov, who prizes free will over monetary gain. If you, the human decision maker, believe the being's initial probabilities place it on S_I , then you can strike a blow for free will by selecting the second box. Of course, you will receive \$0, but you will "feel good." Alternatively, if you believe the being is on (S_{II}) , then you can soundly defeat it: take both boxes and you not only strike a blow for free will but you pick up \$1,001,000.⁶

Can we rig the game against the being? Perhaps. Suppose the being has played the game a number of times against your cohorts, all of whom have, by agreement, selected the second box only. The being, *sans* apparatus, views you as one of your fellows and adopts a high initial probability that

⁵ See e.g., Hilpinen [1968] and Lehrer [1974] and the references cited therein.

⁶ The speed with which you determine that the being is on either (I) or (II) may well be the essence of your free will. See footnote 3.

you too will select the second box only. If the being's initial probability, $P(H_2)$, is sufficiently high to put it on line segment (S_{II}) (by virtue of your cohorts' actions), then you've beaten the being. Regardless of the signal it receives from the apparatus, rule (R_1) has it putting \$M in the second box.

Consider now a being who behaves in accordance with (R_2) . It is not difficult to show that $K- = K+$. Hence, the signal from the apparatus uniquely determines the being's action, so that there are no regions in which the being is open to an "Asimov-strategy," i.e., being beaten in a game rigged as above. If the apparatus is accurate, then the being will be able to discern attempts at deception on your part.

Finally, consider a being who employs (R_3) . This rule, like (R_2) , yields the result that the signal from the apparatus uniquely determines the being's act. As with (R_2) , such a being is not open to an Asimov-strategy. However, if the "apparatus" is more reliable than not on both signals, but not all that reliable, then the human decision maker can gamble on winning \$1,001,000. But what made our being so interesting in the first place was its amazing accuracy. Unless we know that the being has erred, we must behave like the scientist who is presented with an unbroken sequence of successful replications: we must presume that the apparatus is both relatively and absolutely reliable, i.e., that $P(- | H_1)$ and $P(+ | H_2)$ are both very high. Gambling against an (R_3) being would attract only the heartiest of risk lovers.

IV. STATISTICAL PREDICTION

We now turn to prediction techniques available to the being that are based upon variants of Neyman-Pearson and Bayesian statistical procedures. Spielman [1973], in proffering a refutation of the Neyman-Pearson theory of testing, provides acceptance/rejection rules of both the Neyman-Pearson and Bayesian types. Consider the Neyman-Pearson approach first.

Spielman argues that the Neyman-Pearson theory of testing requires that "[one] accept an hypothesis if and only if the reliability index [defined below] for this type of decision is high enough" (Spielman [1973], p. 216). For our problem, the (Spielman) reliability index for a—

$$r- = 1 - \frac{P(- | H_2)p(H_2)}{P(- | H_2)p(H_2) + P(- | H_1)p(H_1)}$$

and that for + is

$$r+ = 1 - \frac{P(+ | H_1)p(H_1)}{P(+ | H_1)p(H_1) + P(+ | H_2)p(H_2)}$$

where $p(H_1)$ and $p(H_2)$ are the proportions of the members of a well-specified reference class who have opted for selecting both boxes and the second box only, respectively. The acceptance rule requires that $r-$ and $r+$ be "high enough." Let $k-$ and $k+$ be the minimally sufficient values. Then the (Spielman) Neyman-Pearson rule is

(R_4) accept H_1 on — iff

$r- > k-$, i.e., iff

$$P(- | H_1) > \frac{k-}{1-k-} \cdot \frac{1-p(H_1)}{p(H_1)} \cdot P(- | H_2)$$

accept H_2 on + iff

$r+ > k+$, i.e., iff

$$P(+ | H_2) > \frac{k+}{1-k+} \cdot \frac{1-p(H_2)}{p(H_2)} \cdot P(+ | H_1)$$

(R_4) is strikingly similar to (R_3) . Indeed, (R_4) implies (R_3) iff $k- = p(H_1)$ and $k+ = p(H_2)$, again where $p(H_1)$ and $p(H_2)$, are expressed as proportions.

Can the (R_4) being be defeated? Consider the \$1,001,000 winning involving, as before, your n cohorts. Suppose they comprise the reference class and all n of them have taken the second box only. Then $p(H_2) = 1.0$ and $p(H_1) = 0$. You plan to take both boxes (H_1) ; suppose the apparatus detects this and reads —. The being will predict correctly, and thereby thwart your efforts, iff $P(- | H_1) > \infty$, since the term $[1-p(H_1)]/p(H_1)$ is $1/0$. By employing the proportions $p(H_1)$ and $p(H_2)$ and allowing the former to take on the value 0, an (R_4) being is indeed open to an Asimov-strategy and the \$1,001,000 result. (Similarly, by allowing $p(H_2)$ to be 0 he is open to the \$0 result.)

Considering somewhat less control over the reference class, suppose k of your n cohorts opt for the second box only. Then $p(H_1) = k/n$, and in order to thwart your attempted \$1,001,000 coup the being must have

$$\begin{aligned} P(- | H_1) &> \frac{k-}{1-k-} \cdot \frac{1-k/n}{k/n} \cdot P(- | H_2) \\ &> \frac{k-}{1-k-} \cdot \frac{(n-k)}{k} \cdot P(- | H_2) \end{aligned}$$

The previous result is obtained here as a limit as $k \rightarrow 0$. For n sufficiently large but finite, k small relative to n , $k-$ high, and a merely reasonable apparatus, i.e., $P(- | H_1)$ not extremely higher

than $P(- | H_2)$, the (R_4) being is open to the \$1,001,000 coup wherein k cohorts take the second box only.

Spielman notes that there is a Bayesian rule, say (R_5) , which is an analogue of (R_4) , and which is indeed similar to (R_4) , "... the difference being that truth-frequencies in a reference class correspond to personal prior probabilities of hypotheses, [and] the reliability index for acceptance of an hypothesis H on outcome E corresponds to the posterior probability of H given outcome E ..." (Spielman [1973], p. 217). We thus have (R_5)

accept H_1 on $-$ iff

$$P(H_1 | -) > k-, \text{ i.e., iff}$$

$$P(- | H_1) > (k-) \frac{P(-)}{P(H_1)}$$

accept H_2 on $+$ iff

$$P(H_2 | +) > k+, \text{ i.e., iff}$$

$$P(+ | H_2) > (k+) \frac{P(+)}{P(H_2)}$$

Two points are of interest. First, if $k-$ and $k+$ are selected in an appropriate fashion, then (R_5) is the acceptance rule AGn, the rule from which (R_1) is taken.⁷ Clearly, (R_5) implies (R_1) iff $k- \geq P(H_2 | -)$ and $k+ \geq P(H_1 | +)$. Since, for a respectable "apparatus" it is reasonable to presume that $P(H_2 | -)$ and $P(H_1 | +)$ are "low" and that for an aware being $k-$ and $k+$ are high, it reasonably follows that if (R_5) is operative, then so is (R_1) . Thus an (R_5) being is open to both the \$1,001,000 and the \$0 outcomes. Note that (R_5) in its basic form, i.e., $P(H_1 | -) > k-$ and $P(H_2 | +) > k+$, is simply a high probability acceptance rule, which is open to the lottery paradox as well as the \$0, and \$1,001,000 results.⁸

V. STRATEGY FOR THE HUMAN DECISION MAKER

How do you, the human decision maker, approach your decision problem? Let us agree that you make your decision after having been observed by the being and its "apparatus" and before the being presents you with the boxes. The state of the being has a five part description. Letting P denote the being's probability measure, its state is specified by $\langle P(- | H_1), P(+ | H_2), P(H_1), \text{rule}, \text{signal} \rangle$.

The values of $P(- | H_1)$, $P(+ | H_2)$ completely determine the being's trust in its "complicated apparatus," that of $P(H_1)$ determines its initial probabilities, the rule specifies its "theory," and the signal (either $-$ or $+$) constitutes its observation. Each state description completely determines a prediction by the being, and hence determines the contents of the second box. You, the human decision maker, will at least have a probability distribution over these five-part states. This distribution, together with your utility structure will determine your optimal action via the usual expected utility approach. Your utility structure deserves comment.

The possible payoffs are \$0, \$1,000, \$M, and \$1,001,000. In every case where the being's decision rule employs $P(H_1)$ and/or $P(H_2)$ explicitly, the potential for an Asimov-strategy exists. In the case of a (R_1) , (R_4) or (R_5) being the \$0 and the \$1,001,000 can be obtained by conspiring with others. You may, therefore, want to employ a conditionalized utility structure wherein, perhaps, the utility of \$0 obtained by defeating the being is greater than that of \$1,000 obtained via less devious procedures. We are tempted to call this an "Asimov-utility structure." Similarly the utility of outwitting the being plus obtaining \$1,001,000 might be so high as to compensate for the low probability associated with achieving this payoff, thereby making the coup the optimal selection. What arises here is a utility structure that must counterbalance the human decision maker's monetary risk aversion and her (or his) proclivity to achieve a coup over seemingly super or beings.

REFERENCES

- ASHBY, W. R. [1963]: *An Introduction to Cybernetics*, J. Wiley, Inc., New York.
- BAR-HILLEL, Maya and MARGALIT, Ashai [1972]: "Newcomb's Paradox Revisited," *British Journal for the Philosophy of Science*, vol. 23, pp. 295-304.
- GARDNER, Martin [1973]: "Free Will Revisited With a Mind-bending Prediction Paradox by William Newcomb," *Scientific American*, vol. 225, pp. 104-108.
- GOOD, I. J. [1971]: "Free Will and the Speed of Computation," *British Journal for the Philosophy of Science*, vol. 22, pp. 48-49.
- HILPINEN, Risto [1968]: "Rules of Acceptance and Inductive Logic," Facs. XXII of *Acta Philosophica Fennica*, North-Holland Publishing Co., Amsterdam.

⁷ Again, for the rule AGn see Hilpinen [1968].

⁸ On the lottery paradox see Hilpinen ([1968], pp. 39-40).

- LEHRER, Keith [1974]: "Belief and Error," in M. Graham and E. Klemke (ed.): *The Ontological Turn*, University of Iowa Press, Iowa City, pp. 216-229.
- LEVI, Isaac [1975]: "Newcomb's Many Problems," *Theory and Decision*, vol. 6, pp. 161-175.
- NOZICK, Robert [1969]: "Newcomb's Problem and Two Principles of Choice," in N. Rescher, *et al.* (eds.): *Essays in Honor of Carl G. Hempel*, D. Reidel Publishing Co., Dordrecht, Holland, pp. 114-146.
- [1974]: "Reflections on Newcomb's Problem: a Prediction and Free-Will Dilemma," *Scientific American*, vol. 230, pp. 102-108.
- SHANNON, C. E. and WEAVER, Warren [1969]: *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.
- SPIELMAN, Stephen [1973]: "A Refutation of the Neyman-Pearson Theory of Testing," *British Journal for the Philosophy of Science*, vol. 24, pp. 201-222.

University of Oklahoma
Central Michigan University
University of Iowa
De Paul University

Received October 7, 1975

BOOKS RECEIVED

- Aagaard-Mogensen, Lars (ed.), *Culture and Art* (Atlantic Highlands, N.J.: Humanities Press, 1976), 211 pp. \$5.00
- Beckner, M. O., *El darwinismo* (Valencia: Revista Teorema, 1976), 70 pp.
- Gould, Carol C. and Marx W. Wartofsky (eds.), *Women and Philosophy* (New York: G. P. Putnam's Sons, 1976), 364 pp. \$8.95
- MacPartland, John, *Conversations with Sheldon* (New York: Philosophical Library, 1976), 90 pp. \$6.00
- Rahman, Fazlur, *The Philosophy of Mullā Ṣadrā* (Albany: State University of New York Press, 1975), 277 pp.
- Rescher, Nicholas, *A Theory of Possibility* (Pittsburgh: University of Pittsburgh Press, 1976), 255 pp.
- Salvucci, Pasquale, *L'uomo di Kant* (Arzalia Editore Urbino, 1975), 637 pp.
- Science and Absolute Values*, Volumes 1 and 2, Proceedings of the Third International Conference on the Unity of the Sciences, London, 1974 (Tarrytown, N.Y.: International Cultural Foundation, 1974), 710 and 1453 pp.

RATIO

edited by STEPHAN KÖRNER

Volume XVIII

December 1976

No. 2

Kant on Existence

J. MICHAEL YOUNG

How to Dispense with Things in Themselves (II)

M. GRAM

Freedom of Indifference

NICHOLAS NATHAN

The Source of the River Time

FERREL CHRISTENSEN

On the Alleged Importance of Moral Rights

WILLIAM NELSON

Mentezin: On the Problem of the Rehabilitation of Criminals

KEVIN MCCORMICK

Reviews:

Gesammelte Werke (Leonard Nelson)

LOTHAR F. NEUMANN

Vom Selbstvertrauen der Vernunft (Leonard Nelson)

RUDOLF MALTER

Contemporary German Philosophy and its Background
(Fritz-Joachim von Rintelen)

RUDOLF TEICHNER

Annual Subscription: £8.00 (inland) £9.00 (overseas) \$24.00 (U.S.A. and Canada)

Published by BASIL BLACKWELL, Oxford

Orders to: Journals Department, Basil Blackwell & Mott Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

Journal for the Theory of Social Behaviour

Edited by P. F. SECORD (Department of Urban Studies, Queens College, New York) and T. MISCHÉL (Department of Philosophy, SUNY at Binghamton, New York) assisted by R. HARRÉ (Oxford University).
Volume 6, No. 2
October 1976

Contents:

STUART ALBERT and SUZANNE KESSLER: Processes for Ending Social Encounters: The Conceptual Archaeology of a Temporal Place
ALLAN R. BUSS and JOSEPH R. ROYCE: Note on the Temporality of Trait Constructs

HARWOOD FISHER: The Language and Logic of Forming an Idea

MICHAEL S. PRICHARD: On Taking Emotions Seriously: A Critique of B. F. Skinner

V. LEE HAMILTON: Role Play and Deception: A Re-examination of the Controversy

Discussions

IAN HARRIS and RICHARD SPANIER: Accountability: Answerability and Liability

CHARLES V. BLATZ: A Reply to Harris and Spanier

JAN NARVESON: Intentional Behaviour and Social Science

Annual Subscription: Institutions £7.50 (inland), £9.00 (overseas) \$20.00 (U.S.A. and Canada); Individuals £4.50 (inland) £5.40 (overseas) \$12.50 (U.S.A. and Canada)

Orders to: JOURNALS DEPARTMENT, BASIL BLACKWELL & MOTT LTD.,
108 Cowley Road, Oxford OX4 1JF, England.

The Monist

*An International Quarterly Journal
of General Philosophical Inquiry*

Founded 1888 by **Edward C. Hegeler**

Editor, Eugene FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor, Ann FREEMAN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate with the Editor in advance for Special Instructions. Papers should be 4000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

SCHEDULED PUBLICATION DATES:

Vol. 60,	No. 1	Jan.,	1977	Bioethics and Social Responsibility
Vol. 60,	No. 2	Apr.,	1977	Philosophy and Religion in the Nineteenth Century
Vol. 60,	No. 3	July,	1977	New Directions in Semantics
Vol. 60,	No. 4	Oct.,	1977	Historicism and Epistemology
Vol. 61,	No. 1	Jan.,	1978	Conceptions of the Self: East and West
Vol. 61,	No. 2	Apr.,	1978	The Philosophy of Thomas Reid
Vol. 61,	No. 3	July,	1978	Nominalism, Past and Present
Vol. 61,	No. 4	Oct.,	1978	The Philosophy and Psychology of Cognition

Editorial Office: Box 1908, Los Gatos, California 95030

Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, \$18.00; Individuals, \$12.00. Single copies: \$4.50. Foreign postage: Add 25 cents to single copy rate or \$1.00 to subscription rate.

Philosophy

The Journal of the Royal Institute of Philosophy

Edited by Renford Bambrough

Vol. 51 No. 198

October 1976

Fifty Years of Philosophy **Gilbert Ryle**

Kant's Kingdom of Ends **Mary A. McCloskey**

Descartes, Beliefs and the Will **Brian Grant**

Who's in Control Here? **Robert V. Hannaford**

The Contemporary Significance of Analogy **Patrick Sherry**

DISCUSSION

NEW BOOKS

BOOKNOTES

Volume 51 (1976) £8.50 (\$23.50 U.S.A. and Canada). Single parts £2.75 (\$7.75 U.S.A. and Canada) each.

CAMBRIDGE UNIVERSITY PRESS

BENTLEY HOUSE, 200 EUSTON ROAD, LONDON NW1 2DB

32 EAST 57th STREET, NEW YORK, N.Y. 10022

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER
in cooperation with a distinguished group of American
and Foreign Scholars

Descriptive, analytic, critical, and historical articles representing the major contemporary trends in philosophy. In addition to phenomenological research, PPR publishes studies in a wide range of areas including ethics and value theory, metaphysics, aesthetics, logic, language, political, social, and religious philosophy, theory of knowledge, and the sciences.

Address all communications to
Philosophy and Phenomenological Research
State University of New York at Buffalo
320 Baldy Hall
Buffalo, New York 14260

Annual subscription rate, \$10.50 for libraries and institutions, \$8.50 for individuals; single copies \$2.75 respectively, plus postage and handling on back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed, stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

THE PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 26, No. 105, OCTOBER 1976

ARTICLES

PLATO'S IMAGE OF IMMORTALITY

Kenneth Dorter

THE ORIGINS OF LOCKE'S DOCTRINE OF PRIMARY AND SECONDARY
QUALITIES

Martha Brandt Bolton

SIDGWICK'S PESSIMISM

J. L. Mackie

REALIZING WHAT'S WHAT

Annette Baier

WHAT IS POETICS?

Stein Haugom Olsen

DISCUSSION

HAMILTON'S THEORY - A CORRECTION

Robert J. Fogelin

CRITICAL STUDY

S. Guttenplan (ed.): *Mind and Language*

Simon Blackburn

BOOK REVIEWS

BOOKS RECEIVED

Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant, Anthony Ellis and Stephen Read.
Published for the Scots Philosophical Club by the University of St Andrews.

Annual Subscriptions: Individuals £3.00 (\$8.50) Institutions £5.00 (\$14.00)

Single Issues £1.00 or \$3.00 (Institutions £1.50 or \$4.00)

for orders placed directly with the Journal Manager, Scottish Academic Press
Ltd., 33 Montgomery St., Edinburgh EH7 5JX, Scotland

(Continued from outside back cover)

COMPLETE LIST OF MONOGRAPH SERIES

- No. 1 (1968) STUDIES IN MORAL PHILOSOPHY. Essays by David Braybrooke, Lawrence L. Haworth, G. P. Henderson, Jesse Kalin, Kai Nielsen, Jerome B. Schneewind, Michael Stocker.
- No. 2 (1968) STUDIES IN LOGICAL THEORY. Essays by James W. Cornman, Montgomery Furth, Alan Hausman and Charles Echelbarger, Jaakko Hintikka, Robert C. Stalnaker, Ted Honderich, Colwyn Williamson.
- No. 3 (1969) STUDIES IN THE PHILOSOPHY OF SCIENCE. Essays by Peter Achinstein, Joseph Agassi, Simon Blackburn, Mario Bunge, Bernard R. Grunstra, Keith Lehrer, D. H. Mellor, Lawrence Sklar, Michael Slote, Stephen Spielman.
- No. 4 (1970) STUDIES IN THE THEORY OF KNOWLEDGE. Essays by Norman Malcolm, W. Donald Oliver, Peter Unger, John L. Pollock, John Knox, Jr., Alan R. White, Frederick Stoutland.
- No. 5 (1971) DESCARTES' PHILOSOPHY OF NATURE. By James Collins.
- No. 6 (1972) STUDIES IN THE PHILOSOPHY OF MIND. Essays by Harold Brown, Haskell Fain and A. Phillips Griffiths, O. H. Green, Donald F. Henze, Moreland Perkins, Alan R. White.
- No. 7 (1973) STUDIES IN ETHICS. Essays by Charles V. Blatz, Clauvia Card, Norman O. Dahl, R. L. Franklin, James D. Wallace, A. D. Woosley.
- No. 8 (1974) STUDIES IN MODALITY. By Nicholas Rescher with the collaboration of Ruth Manor, Arnold vander Nat and Zane Parks.
- No. 9 (1975) STUDIES IN EPISTEMOLOGY. Essays by Dennis M. Ahern, Robert Almeder, Jonathan Bennett, Fred R. Berger, Peter A. French, Margaret Gilbert, David K. Glidden, Mark Pastin, E. E. Sleinis.
- No. 10 (1976) NEGATION AND NON-BEING. By Richard M. Gale.

PRINTED IN ENGLAND

by William Clowes & Sons, Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY
MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

AN INTRODUCTION
TO MODAL LOGIC

by E. J. LEMMON

In collaboration with DANA SCOTT

Edited by KRISTER SEGERBERG

Monograph No. 11

Oxford 1977

This is the latest title in this very impressive monograph series, and we shall be publishing it in March 1977. The text is a collaboration between two well-known logicians, who have produced a lucid introduction to one of the fastest growing branches of modern logic. Not only will the beginner find the book particularly helpful, but it also provides a synthesis of work in the field that will interest advanced students.

To enable you to judge for yourself the range and importance of this series, a complete list of the Monographs is given on the inside back cover. The last issue supplied to institutions on subscription was No. 7 (1973); as from 1977, the price of all monographs to institutions will be \$9.00 (£4.50 to UK subscribers, £5.40 to those overseas). We regret to announce that also from 1977 it will no longer be possible to supply the Monographs free of charge to individual subscribers, and that the cost to them will be \$4.50 (£2.25 to UK subscribers, £2.70 to those overseas).

(Continued on inside back cover)

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrocs
Alvin Plantinga
Jerome Schneewind
George A. Schreier
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 14/NUMBER 2

APRIL, 1977

CONTENTS

- | | | | |
|--|-----|---|-----|
| I. BRUCE RUSSELL: <i>On the Relative Strictness of Negative and Positive Duties</i> | 87 | IV. JOSEPH MARGOLIS: <i>Skepticism, Foundationalism, and Pragmatism</i> | 119 |
| II. EMMETT L. HOLMAN: <i>Sensory Experience, Perceptual Evidence and Conceptual Frameworks</i> | 99 | V. FRANK SNARE: <i>Three Sceptical Theses in Ethics</i> | 129 |
| III. ROBERT MERRIHEW ADAMS: <i>Middle Knowledge and the Problem of Evil</i> | 109 | VI. R. I. SIKORA: <i>The Argument From Analogy Is Not An Argument for Other Minds</i> | 137 |

PAPERS PRESENTED AT THE 51ST MEETING OF THE PACIFIC DIVISION OF THE AMERICAN PHILOSOPHICAL ASSOCIATION (Portland, Oregon, March, 1977)

- | | | | |
|---|-----|---|-----|
| VII. BAS C. VAN FRAASSEN: <i>The Pragmatics of Explanation</i> | 143 | IX. JOHN RAWLS: <i>The Basic Structure as Subject</i> | 159 |
| VIII. JOHN M. COOPER: <i>The Psychology of Justice in Plato</i> | 151 | X. RODERICK M. CHISHOLM: <i>Thought and Its Reference</i> | 167 |
| | | XI. GILBERT HARMAN: <i>How to Use Propositions</i> | 173 |

© AMERICAN PHILOSOPHICAL QUARTERLY, 1977

PUBLISHED IN OXFORD BY BASIL BLACKWELL

ISSN 0003-0481

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes" (short or long). Moreover the editor will only consider papers whose authors certify that while under consideration with us, they will not be submitted elsewhere.

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 12,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The subscription for 1977 is \$28.00 for institutions, \$12.00 to individuals. Single issues will be \$8.00 to all purchasers. For Monograph prices, see below. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The complete series is listed on the back cover. In 1977 these will be available at \$9.00 to institutions and non-subscribers, \$4.50 to individual subscribers. *Please note that it is no longer possible to supply the monographs free of charge to individual subscribers.*



I. ON THE RELATIVE STRICTNESS OF NEGATIVE AND POSITIVE DUTIES

BRUCE RUSSELL

RECENTLY a number of philosophical discussions have appeared in which the question of whether we have a greater duty to refrain from injuring people than to bring them aid is considered. In particular, they have centered about the question whether we have a greater duty to refrain from taking life than we have to save life. In this discussion I will consider in what sense (if any) it could be said that negative duties are stricter than positive duties and, more specifically, whether there is any sense in which the duty to refrain from killing someone is stricter than the duty to save someone.

I will conclude that in an important sense negative duties are not stricter than positive duties. I will consider various examples and arguments that would seem to count against my conclusion. It can be shown, I think, that absurd consequences result if the refraining from injuring/bringing aid distinction is taken to have moral significance in itself. Other principles are available to account for the intuitive judgments we make in those cases generally thought to support the claim that negative duties are stricter than positive ones. Not only will such principles be offered but arguments in support of the claim that negative duties are stricter than positive duties will be criticized and found wanting.

I. THE KILLING/LETTING DIE DISTINCTION

To determine if the difference between killing and letting die is morally relevant, we must carefully distinguish between the two. Jonathan Bennett's argument to the effect that the distinction has no moral relevance has been vitiated by Daniel Dinello's criticism of the way in which he draws it.¹

In this discussion I will adopt a modified version of Dinello's characterization of that distinction.

- (A) x killed y if x caused y 's death by performing movements which affect y 's body such that y dies as a result of these movements.
- (B) x let y die if
 - (a) there are conditions (C) affecting y , such that if they are not altered, y will die.
 - (b) there are certain movements (M) which, if performed by x , will prevent y from dying.
 - (c) x is in a position to perform such movements (i.e., if x tries to perform such movements, he will succeed in performing them and x is able to try to perform them).
 - (d) x fails to perform these movements.
 - (e) y dies as a result of (C).²

Now the above account of letting die does not imply that x is morally responsible if he lets y die. It may not be reasonable to expect x to be aware of (a) or (b). If that were true and if x was not in fact aware of (a) or (b), then he could not be accused of negligence even if (a)–(e) obtained. In addition, even if x were aware of (a) and (b), there could be good moral reasons for him to abstain from performing the movements that would save y . For instance, x might have to take the heart from a healthy z and transplant it in y in order to save him in a case where x would not be morally culpable for failing to do this. Perhaps it is because when we speak of letting die it is often in the context of accusation that some have thought that, in cases like the above, it would be incorrect to say that x

¹ Bennett's discussion is in "Whatever the Consequences," *Analysis*, vol. 26 (1966), pp. 83–102 (esp. pp. 93–98); reprinted in *Moral Problems* (New York, 1971), James Rachels (ed.), pp. 43–66 (esp. pp. 55–60).

Dinello's discussion is in "On Killing and Letting Die," *Analysis*, vol. 31 (1971), pp. 83–86. For Dinello's criticism of the way Bennett draws the killing/letting die distinction, see pp. 84–85.

² This account of letting die differs from Dinello's (p. 85) in that it makes no reference to what x has reason to believe, it contains a parenthetical remark explaining (c) and it includes (e). I am indebted to Fred Berger for discussions which led to this formulation and which concerned points discussed in the following paragraph.

let y die.³ Though the use of the words "letting die" often carries with it the ascription of (moral) responsibility, it need not. We can correctly say, " x let y die" and add, "for good (moral) reasons" (reasons which may include reference to contravening duties or non-culpable ignorance) thereby avoiding any ascription of (moral) responsibility.

(e) is needed since, if (a)–(d) are satisfied, it does not necessarily follow that x let y die. Someone else might intervene and save y while x stands idly by. In that case, x would not let y die. But not only must y die if x lets him die, he must also die as a result of the conditions which endanger his life and about which x is able to do something. If x fails to perform (M) while someone else shoots y in the head and as a result no one is able to save y , then x does not let y die. Given (a)–(d), it is only if y dies as a result of (C) that x can be charged with letting him die.⁴

Given the above account of the distinction between killing and letting die, I will argue that that distinction has no moral significance in itself, i.e., if other things are equal (including considerations about intentions, contravening duties, etc.), there are no cases in which the killing/letting die distinction has moral significance.

If the above account is broadened by replacing "killed," "dies," etc. by appropriate versions of "harmed," "is harmed," etc., an account of the distinction between harming and allowing harm will result.

II. THE STRICTNESS OF DUTIES

There are (at least) three senses in which one duty can be said to be stricter than another.

- (1) *Moral Culpability*: Duty S is stricter than duty T if, and only if, other things being equal, one is more culpable for not having performed S than for not having performed T .

For example, the duty to save a life is stricter in this sense than the duty to relieve discomfort and

the duty to refrain from taking life is stricter than the duty to refrain from causing discomfort.

- (2) *Utilitarian Overrides*: Duty S is stricter than duty T if, and only if, other things being equal, a greater amount of non-moral good (e.g., pleasure, freedom from pain and suffering, etc.) is required to justify non-compliance with S than with T .

(1) and (2) are not equivalent for one might hold that of two duties, S and T , no amount of non-moral good could justify non-compliance with either of the duties. Thus it would not be the case that a greater amount of non-moral good would be required to justify non-compliance with S than with T (or vice versa) in sense (2). Nevertheless, one might wish to maintain that S was a stricter duty than T in sense (1). Some would hold that if S were the duty to refrain from killing someone in order to save others and T were the duty to refrain from torturing someone to prevent someone else from torturing others, then S would be stricter than T in sense (1) but not in sense (2). This might also be held to be the case if S and T were not both negative duties but if S were a negative duty and T a positive duty. For instance, let S remain as above and let T be the duty to save life.

- (3) *Perfect vs. Imperfect Duties*: Duty S is stricter than duty T if, and only if, other things being equal, the amount of sacrifice owed to relevant individuals in the performance of S is greater than that for T .

To say that the amount of sacrifice owed to relevant individuals is greater in the case of one duty than in the case of another means either that one would be more culpable for non-performance in the one case than in the other or that a greater amount of non-moral good would be required to override the one duty than the other. (3) applies the senses of strictness defined in (1) and (2) to considerations about the amount of sacrifice owed to individuals. The duty to save life might be just as strict in sense (1) or (2) as the duty to refrain from taking

³ David H. Sanford has claimed that a doctor cannot be accused of letting someone die if he fails to kill a healthy person to obtain a needed heart; cp. "Killing and Letting Die," p. 3, delivered at the Western Division Meeting of the American Philosophical Association, April 30, 1976. John Casey has tried to present reasons why it would be improper to say that the doctor lets the patient die in cases of this type; cp. "Actions and Consequences" in *Morality and Moral Reasoning* (London, 1971), pp. 154–205, esp. Parts II and III (pp. 165–89). All Casey has shown, I believe, is that because of conversational or contextual implicature it is misleading or even improper to claim that the doctor lets the patient die in the relevant circumstances. But this does not show that it is incorrect to claim that, nor need such a claim be misleading if the proper qualifications are made (e.g., in terms of good moral reasons).

⁴ Sanford (*op cit.*, p. 4) gives reasons for requiring that y die (if x lets him die) but fails to note that y must die as a result of the relevant causes.

life when the first kind of duty is taken to be owed to people in general or to some group of people, while the second duty might be stricter than the first if the first is taken to be owed to specific individuals. Why this should be so will be discussed below (VB, 3).

III. CASES IN SUPPORT OF THE CLAIM THAT THE KILLING/LETTING DIE DISTINCTION HAS MORAL SIGNIFICANCE IN ITSELF

Dinello gives a case which lends support to the claim that the killing/letting die distinction (when correctly drawn) has moral significance.

- C_1 : Consider the following example: Jones and Smith are in a hospital. Jones cannot live longer than two hours unless he gets a heart transplant. Smith, who had one kidney removed, is dying of an infection in the other kidney. If he does not get a kidney transplant, he will die in about four hours. When Jones dies, his one good kidney can be transplanted to Smith, or Smith could be killed and his heart transplanted to Jones. Circumstances are such that there are no other hearts or kidneys available within the time necessary to save either one. Further, the consequences of either alternative are approximately equivalent, that is, heart transplants have been perfected, both have a wife and no children, *etc.*⁵

Dinello then goes on to claim that there is a greater duty to refrain from killing Smith than there is to save Jones (by killing Smith). This is a difficult case for anyone who would hold that in itself the killing/letting die distinction has no moral significance. In order to maintain that position and accept Dinello's judgment on (C_1), some morally relevant aspect of this case must be found which is not essentially related to the killing/letting die distinction but that is covertly introduced along with the other aspects of the case. Another alternative would be to present reasons to show that Dinello's intuitions are not well-founded here.

Richard L. Trammell presents another difficult case for those who deny the killing/letting die distinction has moral significance.

- C_2 : If someone threatened to steal \$1000 from a person if he did not take a gun and shoot a stranger between the eyes, it would be very wrong for him to kill the stranger to save his \$1000. But if someone asked from that person \$1000 to save a stranger, it would seem that his obligation to grant this request would not be as great as his obligation to refuse the first demand—even if he had good reason for believing that without his \$1000 the stranger would certainly die.⁶

Philippa Foot presents a number of examples that are meant to support the claim that in general we have a greater duty to refrain from injuring than to bring aid. I will give three of her examples because I think they contain important differences.

- C_3 : Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed.⁷

To make this case symmetrical with others she presents, Foot later alters it slightly so that the rioters threaten to take the lives of five hostages rather than the lives of certain members of some community. A family of similar cases can be generated from this example. All of them will involve the choice between helping some by *injuring* or *killing* others or letting those in need of help be injured or killed by refusing to kill or injure others.

- C_4 : A person can help others by deliberately letting someone die (e.g., by letting a starving beggar die) and then using that person's vital bodily parts (heart, kidneys, etc.) in order to save others.⁸

(C_4) differs from (C_3) in that *both* alternatives involve *letting* someone *die*. In (C_3) *one* of the alternatives involves *killing* someone to aid others; the other, letting some persons die.

- C_5 : Suppose, for instance, that there are five patients in a hospital whose lives could be

⁵ Dinello, *op cit.*, pp. 85–86.

⁶ Richard L. Trammell, "Saving Life and Taking Life," *The Journal of Philosophy*, vol. 72 (1975), p. 131.

⁷ Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," from the *Oxford Review*, no. 5 (1967). Reprinted in *Moral Problems*, J. Rachel (ed.), (*op cit.*), p. 32. All references to this article are to the reprint.

⁸ Foot, *ibid.*, p. 38.

saved by the manufacture of a certain gas, but that this inevitably releases lethal fumes into the room of another patient whom for some reason we are unable to move.⁹

This case is significant because Foot thinks that defenders of the doctrine of double effect must conclude that it is morally permissible to produce the gas which she claims runs counter to our moral intuitions. Defenders of the moral significance of the killing/letting die distinction have no difficulty with this case since producing the gas would amount to *killing* the immobile patient while not producing it would only be *letting* the other patient *die*.¹⁰

In an essay entitled "The Survival Lottery," John Harris defends a kind of lottery where people are chosen at random to supply vital organs "whenever doctors have two or more dying patients who could be saved by transplants, and no suitable organs have come to hand through 'natural' deaths."¹¹ We will call this case (C_6) where the vital organs of some healthy person are taken in order to save others. Harris argues that such an action is morally permissible while indicating that he believes it is contrary to our moral intuitions to claim such.¹² Because this example has significant features in common with (C_1 – C_5), it must be placed in the same class with them whatever the final ruling might be as to the moral permissibility of the envisioned actions.

IV. CASES IN SUPPORT OF THE CLAIM THAT THE KILLING/LETTING DIE DISTINCTION HAS NO MORAL SIGNIFICANCE IN ITSELF

Trammell cites two examples given in different papers by Michael Tooley that seem to count against those who hold that the killing/letting die distinction has moral significance in itself. I will give Tooley's case of a diabolical-machine and

alter Trammell's (C_2) so that it appears to count against those who maintain that the killing/letting die distinction has moral significance.

D_1 : Imagine a machine which contains two children, John and Mary. If one pushes a button, John will be killed, but Mary will emerge unharmed. If one does not push the button, John will emerge unharmed, but Mary will be killed. In the first case one kills John, while in the second case one merely lets Mary die. Does one really wish to say that the action of intentionally refraining from pushing the button is morally preferable to the action of pushing it, even though exactly one person perishes in either case?¹³

D_2 : If someone threatened to steal \$1 from a person if he did not take a gun and shoot a stranger between the eyes, it would be very wrong for him to kill the stranger to save his \$1. If someone asked from a person \$1 to save a stranger, it would seem that his obligation to grant this request would be just as great as his obligation to refuse the first demand—given that he had good reason for believing that without his \$1 the stranger would certainly die.

I have modified Foot's (C_3) along the lines of her (C_4) in order to show that there is not always a greater duty to refrain from injuring than to bring aid.

D_3 : Suppose that the innocent man the judge has in his custody is a diabetic. Now he can obtain the release of the five hostages by (i) *letting* the diabetic die through refusing him insulin or by (ii) *killing* him through hanging, shooting, etc. On the other hand, he can (iii) refuse to either

⁹ Foot, *op cit.*, pp. 38–39.

¹⁰ James Hanink fails to deal with this case in his article in defense of the doctrine of double effect. To be consistent he would have to treat it as he does the case where a fat man is trapped (face pointing outward) in the only exit from a cave. We can suppose just one person (instead of several) is trapped inside the cave, which is rapidly filling with water, and that he will drown if a stick of dynamite is not discharged to blow the fat man out of the exit. For Hanink, blowing the man free does not seem intuitively wrong. Given this, I suspect his intuitions in the gas/lethal fumes case would be that it was morally permissible to produce the gas. To judge the case in that way would seem to be counter-intuitive.

See James Hanink, "Some Light on Double Effect," *Analysis*, vol. 35 (1975), pp. 147–151. See p. 149 for the case of the trapped fat man, a case that also appears in Foot, *op. cit.*, pp. 31, 39–40. For criticism of Hanink's position, see R. A. Duff, "Absolute Principles and Double Effect," *Analysis*, vol. 36 (1976), esp. pp. 68–69.

¹¹ John Harris, "The Survival Lottery," *Philosophy*, vol. 50 (1975), p. 83.

¹² Harris, *ibid.*, pp. 81, 87.

¹³ Cited in Trammell, *op. cit.*, p. 136. Taken from Michael Tooley's "Abortion and Infanticide Revisited," delivered at the American Philosophical Association Meeting, Eastern Division, held in Atlanta, Georgia, December 27–29, 1973.

kill the diabetic or let him die and thereby let the hostages be killed.

If the killing/letting die distinction is in itself morally significant, then the judge would have a stricter duty (in one of the three senses of "strictness") to let the diabetic die than to kill him. But what difference could it make whether the judge killed the innocent man or let him die by refusing him insulin, other things being equal? If the distinction is taken to have moral significance in itself, when applied to this modified judge-rioters case it leads to absurd consequences.

V. ARGUMENTS IN FAVOR OF THE MORAL SIGNIFICANCE OF THE KILLING/LETTING DIE DISTINCTION

A. Harman

In a discussion whose purpose is to show that "morality arises when a group of people reach an implicit agreement or come to a tacit understanding about their relations with one another," Gilbert Harman attempts to explain what he assumes to be the case, viz., that "our morality takes it to be worse to harm someone than to refuse to help someone."¹⁴ His hypothesis is that moral agreement results from bargaining with conditional intentions. Applying these considerations to the harming/helping distinction, Harman reasons that everyone (the rich, the poor, the strong, the weak) would agree to avoid harming one another since that would benefit all. However, "the rich and the strong would not benefit from an arrangement whereby everyone would try to do as much as possible to help those in need."¹⁵ Given this situation, the weak and the poor hold out for the strong principle of mutual aid while the rich and the strong favor only a very weak one, perhaps one that requires everyone to do something (perhaps very minimal) to help those in need.

Harman suggests that the result is a compromise between the strong version of the principle of mutual assistance offered by the weak and the poor and the weak version offered by the rich and the strong. On the other hand, the principle of refraining from harming has unanimous support and is thus very strong. This account is meant to explain why we (purportedly) take it to be worse to harm someone than to refuse him aid. If Harman's

bargaining hypothesis about how we come to adopt certain moral attitudes were correct, we might use it to support the claim that any group of people that contained rich and strong, weak and poor, *ought* to bargain for a stronger duty to refrain from harming than to bring aid since that would be the best bargain to strike, even if they did not in fact realize it.

But the bargaining procedure Harman proposes is insufficient for determining *moral* principles. We can imagine agreement on all sorts of principles of etiquette being reached as a result of the kind of bargaining process Harman offers and other clearly immoral bargains being struck on the basis of race, sex, religion, etc. To rule out agreements based on bargains of the latter sort, Harman would have to introduce something comparable to Rawls' veil of ignorance.¹⁶ When that is done, it is not clear that the parties would see the duty to refrain from injuring as stricter than the duty to bring aid since from behind the veil of ignorance they would not be able to tell whether, as people in ordinary life, they would need aid or not. As it stands, Harman's account of the bargaining process is not an adequate account of a procedure for determining moral principles since it makes no provision for fairness or impartiality between the bargaining parties. Thus, his conclusion concerning the relative strictness of negative and positive duties is not well founded.

B. Trammell

In his discussion, Richard L. Trammell offers three reasons why he thinks there is a greater obligation not to kill than to save life.

(1) *Optionality*

Trammell notes that certain kinds of action destroy the possibility of other people's intervening to prevent evil from occurring while others do not. He then says (correctly, I believe) that killing someone closes off everyone's options to realize a good (viz., to keep the dead person alive) but that passing a dying person by, when one can help to prevent his death, leaves open the possibility of others coming to his assistance. Given that it is better to adopt an alternative that has less chance than some other of producing the same amount of non-moral evil (e.g., pain, physical disability,

¹⁴ Gilbert Harman, "Moral Relativism Defended," *The Philosophical Review*, vol. 84 (1975), pp. 3, 13.

¹⁵ Harman, *ibid.*, p. 12.

¹⁶ See John Rawls' *A Theory of Justice* (Cambridge, 1971), p. 139 where he points out bargaining in "the usual sense" is ruled out by the veil of ignorance.

death, etc.), it would be better to pass a dying person by than to kill someone (other things being equal, of course).

While this conclusion holds, I think, when killing is compared with passing a dying person by, it does not hold when killing is compared with letting someone die. Whether a person adopts a plan to kill someone or to let him die, the death is logically necessary given successful execution of the relevant plan. On the other hand, if the relevant omission involves passing by a dying person without offering assistance and the relevant act, shooting someone in the head at close range, then in neither case is the death of that person logically necessary. It might even be the case that the probability of death given the relevant omission is greater than that given the relevant act.¹⁷ If one gives equivalent descriptions of corresponding negative and positive duties, then optionality either becomes an accidental feature or necessarily not a feature of non-compliance with either kind of duty. So any moral significance optionality may have cannot be used to account for the moral significance (if any) of the refraining from injuring/bringing aid distinction.

(2) Responsibility

An argument can be constructed from Trammell's discussion of responsibility that seems to support the claim that there is a greater duty to refrain from injuring than to bring aid.¹⁸ It rests on a premise that connects non-moral responsibility for states of affairs with the strictness of moral obligations. It seems that for Trammell a person is non-morally responsible for a state of affairs if, and only if, he has either caused or somehow brought about that state of affairs. He argues that if a person is non-morally responsible for a state of affairs that injures or endangers someone else, that person is under a greater obligation to aid the injured or endangered person than are those who are not so responsible. He claims that the obligation to bring aid that the person who is responsible has incurred is just as great (or nearly as great) as his obligation to refrain from causing the injury or danger in the first place.

If we let X = the strictness of the duty to bring aid on the part of those who are *not* non-morally responsible for some injury or danger, Y = the strictness of the duty to bring aid on the part of those who *are* non-morally responsible for some injury or danger, and Z = the strictness of the obligation not to injure or endanger in the first place, then the argument has the following structure:

$$(1) X < Y$$

$$(2) Y = Z$$

$$(3) X < Z$$

From this one can conclude that, in general, (except where one has caused the injury) the duty to bring aid is not as strict as the duty to refrain from injuring.

This argument fails on two counts. First, (1) is not obviously true. If someone accidentally (and non-negligently) nudged another person (a non-swimmer) into the water, he would be non-morally responsible for the resulting predicament. But if both he and another man knowingly let the non-swimmer drown when they could easily have saved him, it would be beside the point should the other man plead for a lighter sentence on the grounds that *he* did not accidentally nudge the drowned man, i.e., on the grounds that he was not non-morally responsible for the drowned man's predicament. If the cause that sent the non-swimmer into the water was truly accidental (and non-negligent), then what moral difference could it make whether the cause was an unintended nudge rather than merely a slippery deck? To call the act accidental and non-negligent is to imply that the relevant person is completely absolved of moral responsibility for the resulting situation. If the person has no moral responsibility for the situation, there would seem to be no grounds for assigning him a greater duty for changing it than others similarly situated.

Let us now suppose that someone intentionally or negligently nudged a non-swimmer into the water. Would he have a greater duty to bring aid to the non-swimmer than an innocent bystander? I

¹⁷ Jonathan Bennett has noted the moral significance of inevitabilities and expected inevitabilities. He has argued that the greater inevitability of harm does not *necessarily* attach to violations of negative rather than positive duties (Bennett, *op. cit.*).

¹⁸ I say "constructed" because the argument as it stands is confused. It is not always clear whether Trammell is using "responsible" in a moral or a non-moral sense. Also Trammell's conclusion that "a person is not necessarily responsible for someone else's needing to be saved; but he is responsible for the life of anyone he kills" (p. 136) is either false, irrelevant or suffers from ambiguity depending respectively on whether "responsible" is being used morally or non-morally throughout or morally in one place and non-morally in the other.

think that he would, at least if strictness is understood in the third sense given above. In order to right the wrong he has done, this man might be required to jump into shark-infested waters to save the drowning non-swimmer while the innocent bystander would not be morally blameworthy should he refrain from taking the plunge under these circumstances. But in this case it is plausible to maintain $Y > Z$, i.e., that (2) is false. In general, no one has a duty to refrain from injuring someone where that duty is so great that it would require one to endanger oneself in a way comparable to that which would result from jumping into shark-infested waters.¹⁹ But if initially a person cannot be required to sacrifice so much in order to be morally blameless but can be required to sacrifice that much if he is to be blameless after intentionally (or negligently) injuring or endangering another, it makes sense to say that the duty to bring aid under those circumstances is stricter than the initial duty to refrain from injuring or endangering. Thus $Y > Z$.

In summary, if a person is non-morally responsible but *not* morally responsible for the injury or danger to another, then (1) is false. On the other hand, if a person is non-morally responsible *and* morally responsible for the injury or danger to another, then (2) is false. Thus, in any case, the argument that I have presented and that can be gleaned from Trammell's discussion fails to support the desired conclusion.

(3) *Dischargeability of a Duty*

Trammell rightly points out that given the world as it is, it is impossible for one person "to aid everyone who needs help."²⁰ Yet it is infrequent that we need expend great effort or incur great risk to avoid harming people. His point seems to be that the continuous and substantial effort required, if it were a duty to aid everyone who needs help, would be too much to expect of people. It would be unreasonable to require people to forgo their own interests so completely. However, the sacrifice needed to fulfill the negative duty of refraining from killing is generally slight and easily discharged. Thus it is reasonable to require people to refrain from killing others but not reasonable to require them to save everyone.

One might well grant Trammell his point if the positive duty to bring aid is understood as a duty "to aid *everyone* who needs help" or even to aid everyone whose life can be saved. But suppose there are duties to bring aid that are not so general, duties whose formulation contains reference to limits upon the effort or risk that can be reasonably required of those who are to fulfill them. For instance, one might claim that there was a duty to save those whose life was in danger provided one did not have to risk one's own basic interests. We must now see why the principle that limits the amount of sacrifice required to perform what would otherwise be morally obligatory acts leads to the conclusion that negative duties are, in one sense, no stricter than positive duties.²¹

Suppose there is a person who is in need of aid and that we can bring aid to him. Suppose we can assign a number that represents the amount of effort or risk we would have to assume in order to give the person the aid he needs. Let this number be represented by 'e'. Now also suppose there are $(n-1)$ people who are in relevantly similar situations to the person who needs our help. Next, suppose that 'B' represents the greatest sacrifice or effort we can offer without endangering our basic interests. Now given the formal principle of justice that enjoins treating relevantly similar cases in similar ways, we would have the same degree of duty to all the n people. But given that in general we cannot be required to forgo our basic interests, it follows that $n \times e \leq B$. Whenever the effort or risk (e) and/or the number of similarly situated people (n) becomes so great that if we had a duty to any one individual, our obligations would become so great as to jeopardize our basic interests, then, barring overwhelming utilitarian considerations, we have no obligation to any particular individual in the relevant circumstances. So if there are a large number of needy people similarly situated, we will not owe each individual very much effort or be required to incur anything but minimal risks. If we did owe any one individual very much, then given the formal principle of justice and the large number of similarly situated people, we would be required to perform what would more properly be called supererogatory acts.²²

¹⁹ See the next section on the dischargeability of duty where legitimate limitations on moral obligations are discussed.

²⁰ Trammell, *op. cit.*, p. 133.

²¹ Cp. Rawls' account of supererogatory acts as those that would normally be required "if not for the loss or risk involved for the agent himself," *A Theory of Justice*, *op. cit.*, p. 117. Cp., also, p. 439.

²² I believe it is these three factors, viz., the formal principle of justice, a large number of needy persons similarly situated and the principle of legitimate limitations on moral obligations, that led Kant and Mill to distinguish between perfect and imperfect

But these considerations do not show that the refraining from injuring/bringing aid distinction has moral significance *in itself*. On the contrary, one would think that in certain situations where the sacrifice required to aid someone was slight and/or the number of relevantly similar cases was small, the duty to bring aid in these cases would be just as strict (in the third sense of "strictness") as the duty to refrain from injuring. (Think of a case where all one need do to save a drowning man is throw him a life preserver.) Whether the effort or risk required to bring aid in a particular case is great or small is not of paramount importance.²³ What is important is the total amount of effort or risk required if one is to treat all relevantly similar cases in the same manner. The fact that we are generally required to expend greater effort and incur greater risks to avoid harming any particular individual than we are to bring him aid can be explained on the basis of the fact that there are relatively few individuals situated in such a way that we must go to great lengths to avoid harming them while there are relatively many individuals who would benefit from even slight effort on our part to bring them aid.²⁴

The apparent difference between (C_2) and (D_2) can now be accounted for. Depending upon the amount of relevantly situated people in the world and depending upon one's own financial resources, it might or might not be a stricter duty (in sense 3) to refrain from injuring than to bring aid in each of the respective cases. The reason we tend to think that it would not be a stricter duty in (D_2) is because \$1 represents a rather small drain on the financial resources of most people. Given that there are relatively few people in the world whose lives could be saved by the expenditure of \$1, the total burden of treating relevantly similar cases in the same way (viz., by giving a dollar) would not ex-

ceed reasonable limits. But \$1000 represents a considerable drain on the financial resources of most people. Given that there are relatively many people in the world whose lives could be saved by the expenditure of \$1000, the total burden of treating relevantly similar cases in the same way (viz., by giving \$1000) would exceed reasonable limits—even for a Rockefeller. However, for many people a onetime expenditure of \$1 (or even of \$1000) would not require them to forgo their basic interests. And, for the most part, one need not risk \$1 or \$1000 to avoid killing someone. The apparent difference between (D_2) and (C_2) and between the alternatives in (C_2) and (D_2) are a result of important contingencies in the world. But these contingencies are not essentially attached to the refraining from injuring/bringing aid distinction.

VI. ASYMMETRIES

So far I have criticized arguments offered in support of the claim that negative duties are stricter than positive ones. I have concluded that as long as the fulfillment of positive duties does not jeopardize a person's basic interests, his duty to fulfill them is as strict as his duty to fulfill negative duties, i.e., just as much effort and risk is required in the one case as in the other. But what about cases where to bring aid would require someone to forgo his basic interests while failure to do so would have the consequence that others would be required to forgo theirs? All the cases listed above (with perhaps the exception of C_2 and D_2) appear to be symmetrical with respect to the neglect of basic interests since all the alternatives within each particular case require that someone give up his basic interests. In these cases there is no need to consider how many people are similarly situated

duties. It is not that people are not morally required to expend as much effort or incur as much risk to help the needy as to avoid injuring anyone. It is rather that the amount of effort can be distributed with a certain amount of leeway when it comes to aiding the needy, given that to try to aid everyone would lead one beyond the bounds of moral obligation. See *Utilitarianism* (New York, 1948), pp. 61–62, and for Kant's distinction between perfect and imperfect duties see *Groundwork of the Metaphysics of Morals*, tr. by H. J. Paton (New York, 1964), p. 89n.

²³ Trammell allows (p. 133) that where only minimal effort is required either to bring aid or to avoid injuring there may be no morally relevant difference between violating the respective duties. But whether there is a morally relevant difference or not can only be determined by looking beyond the particular case to other relevantly similar ones.

²⁴ These considerations can form the basis of an argument for moving beyond what Robert Nozick has called "the minimal state." If individuals have positive as well as negative rights (or the correlative duties), then there is no reason the state should not protect these as well. Since no individual can be required to aid all the needy people in the world, each needy person cannot have a right against any particular individual among the better-off to meet all his needs. But each needy person might have a right to have the better-off as a group meet his needs, i.e., each needy person would be entitled to have his needs met by the people who comprise the better-off class, while, of course, there ought to be a fair distribution of the total burden among the members of the better-off group. See Nozick's *Anarchy, State, and Utopia* (New York, 1974), esp. pp. ix and 10 where his narrow, Lockean conception of rights is offered.

and no need to appeal to the formal principle of justice to determine whether, in the relevant circumstances, someone's basic interests will be jeopardized. The adoption of the relevant alternative will by itself result in the death of some person and thus will require him to forgo his basic interests (if he has any).

To account for our considered judgments, which tell us that it is not morally indifferent which alternative is adopted in such cases, some have appealed to the principle that, other things being equal, it is worse (morally) to kill than to let die. This principle accounts nicely for our judgments in certain hostage (e.g., C_3) and "hospital" (e.g., C_1 , C_5 , C_6) cases. But there are other "hospital" (e.g., C_4) and hostage (e.g., D_3 (i) and (iii)) cases where the principle gives no guidance since both alternatives involve letting someone die. In other cases, the principle seems to give us improper guidance since D_3 (i) (an alternative that involves letting die) would be preferable to D_3 (ii) (an alternative that involves killing) by that principle. In a case similar to (D_3) but involving a runaway railway train, if the driver has to *do* something (e.g., turn a wheel, pull a lever, etc.) to steer down track A upon which one person is tied, but can remain on track B , where one person is also tied, *without doing* anything (e.g., by keeping his hands off a wheel, a lever, etc.), then by the principle under consideration he would have a greater duty to stay on B than to steer down A (other things being equal). However, I think most of us would judge that it is morally indifferent which alternative the driver adopts and that D_3 (i) deserves the same moral judgment as D_3 (ii), whatever that may be. These examples show, I think, that the killing/letting die distinction has no moral significance in itself.

However, there is another asymmetry that exists between the alternatives in all of the examples which, I believe, can account for the moral judgments we make in these cases. The asymmetry is that the death of some person will contribute to the well-being of others while the death of those others will not contribute to the well-being of the relevant person. This is easily seen in the hostage cases where it is assumed that the death of the man in the judge's custody is needed to save the hostages but the reverse is not true.²⁵ In these cases it seems appropriate to say that killing the man (or letting him die) is a means to saving the hostages.

But in the hospital cases (especially C_5) it is not so clear that the death of the relevant patient is a means to saving the others. In (C_1 , C_5 and C_6) all one needs are the vital organs of some donor, not his death. In (C_3) all one needs is to produce a gas, the production of which happens to release lethal fumes into the room of an immobile patient, not the death of that patient. The difference between the relevant alternatives in the hostage case and those in the hospital cases is that in the former the death of some person is needed to save others while in the latter all that is needed is some act (or omission) that will inevitably lead to someone's death if performed (or permitted). To say the death, the acts or the omissions are needed means that it is only through them that there is any likelihood the relevant persons can be saved. In none of the cases is the death of someone *logically* necessary to save those whose lives are in jeopardy. The rioters could all faint and the hostages walk away unharmed, and it is possible for the needy patients to make miraculous recoveries. But in the hostage case it is a logically necessary component of a plan that must (in a non-logical sense) be adopted if the relevant people are to be saved. In the hospital cases not only is the death of someone not logically necessary to save those in need, it is not even a logically necessary component of the plan needed (in a non-logical sense) to save the dying patients. To say that something is a logically necessary component of some plan means that any proper description of that plan must be formulated in terms of that something.

Defenders of the doctrine of double effect try to exploit the kinds of differences that appear between the hostage and the hospital cases. When they do so, they rule cases like the hostage ones to be morally impermissible on the grounds that the death (or harm) to the relevant persons would be directly intended while ruling that some cases that have features in common with the hospital ones are morally permissible on the grounds that in those cases death is foreseen but not directly intended. While the doctrine of double effect can account for some of our considered judgments,²⁶ it can be shown to lead to absurd consequences. We can show this by modifying the hostage case. Suppose the rioters condition the release of the hostages upon the removal of all the vital organs from the body of the relevant person, rather than upon his death. Given the choice, would it be morally

²⁵ Foot, *op. cit.*, p. 33.

²⁶ Foot, *op. cit.*, 32-34.

preferable to meet that demand rather than the one requiring the man's death? I do not see how it would or how one could even bring about the death without adopting some plan comparable to one that calls for the removal of vital organs. This modification of the hostage case suggests that there is no morally relevant difference between adopting a plan where death is logically necessary and one where it is only physically inevitable, given the successful execution of the relevant plan.

We are now in a position to formulate a principle which, I think, can account for most of our considered judgments in the relevant conflict of interest cases. The principle is that, other things being equal, it is worse (morally) to adopt a plan where the death of someone is needed (in a non-logical sense) to save another than it is to let that other die regardless of whether (a) the death is brought about through an action or an omission and whether (b) it is logically necessary or only physically inevitable, given the successful execution of the plan. While it may be the case that only where death is a logically necessary component of the relevant plan is it proper to call it a means to saving those in need,²⁷ given the examples offered above, I think it is proper to say that even when death is not a logically necessary component it can nevertheless be treated *as if* it were a means to the appropriate end. In any case, it seems proper to say that *the people*, for whom death is a certain outcome upon successful execution of the relevant plan, are being used as means (only), even if their deaths are not strictly speaking means to the relevant end. To adopt a plan where someone is killed or let die for the sake of others is to treat the relevant person as a means only. To use people in that way is an affront to their worth and dignity in a way that simply letting someone die is not, given there are no means available (short of bringing about the death of another) for saving that person. This is an important consequence since it shows that not only does the principle I have offered account for most of our considered judgments, but that it also connects with important moral concepts, concepts concerning the worth and dignity of persons.

There are several ways in which the principle can be broadened. First, a more general principle concerning harming and allowing harm instead of killing and letting die can be formulated. Second,

it is not really important whether the death of someone is needed to save another or not. There would not be much point in adopting a plan which called for the death of someone as a "means" to saving others if there were other plans available that would accomplish the same end without that cost (assuming complete knowledge of alternatives). Nevertheless, should a person adopt such a plan in those circumstances, it would be worse than letting the relevant person die, though, of course, the best alternative would be to adopt the plan that promised to save the dying person at little or no cost to others. We need not require that the relevant plan be the only one with any likelihood of saving the relevant people. One could be led to that conclusion because that feature happens to belong to the relevant alternatives in most conflict of interest cases involving the lives of various people. Given these considerations, the following more general formulation of the principle seems more appropriate: other things being equal, it is worse (morally) to adopt a plan which calls for harming some in order to aid others than it is to let those others be harmed regardless of whether (a) the harm to some is brought about through an action or an omission and whether (b) the harm is logically necessary or only physically inevitable, given the successful execution of the plan.

All that remains is to check the application of the principle against our considered judgments in the diabolical-machine case (D_1) and in the double-transplant case (C_1). As it stands, (D_1) is inconclusive. What our judgments would be and what the application of the above principle would yield depends on the inner workings of the machine. If pushing the button releases lethal fumes into John's chamber while not pushing it allows an already dying Mary to succumb, then I think our considered judgments would be the same as those in (C_5) and application of the principle would yield the same results. On the other hand, if pushing the button causes a railway train to steer down track *A* upon which John is tied while not pushing it allows it to proceed down track *B* upon which Mary is tied, then I think our intuitions would tell us it makes no difference which alternative is adopted. Again, application of the principle prohibiting the use of people as a means would yield the same conclusion.

²⁷ Warren Quinn criticized a point in an earlier paper of mine on the grounds that it is not strictly proper to call something a means to a given end just because, in the circumstances, it is an empirically inevitable consequence of something which is strictly speaking a means to that end.

(C_1) presents a slightly more difficult case since it seems that the only thing that could account for our judgment that Jones ought to be allowed to die rather than killing Smith is that the one alternative involves letting die while the other involves killing. But I have given reasons to discount the moral relevance of the killing/letting die distinction and have explicitly discounted it in the formulation of my principle. On closer examination, though, we find that there is a relevant asymmetry in this case since it would not be proper to say Jones was let die *in order to* save Smith while it would be proper to say Smith was killed in order to save Jones should the respective alternatives be adopted. The principle I have offered says, in its more specific formulation, that it is worse to kill or let die *in order to* save others than it is to simply let those

others die, other things being equal. Thus, there is an asymmetry between the alternatives in (C_1) that allows my principle to take hold and yield judgments that square with our moral intuitions.

This discussion is now complete. I have tried to show that positive duties are as strict as negative duties as long as complete fulfillment of them does not infringe upon a person's basic interests and that there is a principle available which will offer guidance where there is a conflict of basic interests. This principle has none of the disadvantages of either the doctrine of double effect or of the killing/letting die distinction. Besides yielding judgments that square with our moral intuitions, it is essentially connected to notions of personal dignity and worth. As such, it should be doubly difficult to unseat.²⁸

University of California, Davis

Received February 5, 1976

²⁸ I would like to express my indebtedness to Fred Berger, Warren Quinn, Tom Shillock and Bonnie Steinbock whose comments on the earlier paper I delivered at the Pacific Division Meeting of the APA have helped to make this discussion better than it otherwise would have been. Special thanks are due my ex-colleague Murdith McLean. It was he who first suggested to me that the notion of using people as means was important in the relevant conflict of interest cases. I have benefited in untold ways from the discussions we have had on this and other philosophical topics. Finally, I would like to thank Hugh Fitzpatrick for helping me to pare down to size a longer version of this paper. Of course, none of these people are responsible for any errors that may remain.

II. SENSORY EXPERIENCE, PERCEPTUAL EVIDENCE AND CONCEPTUAL FRAMEWORKS

EMMETT L. HOLMAN

MUCH of the action in the theory of perception over the past couple of decades has centered around the pros and cons of various versions of direct realism. By "direct realism" I here mean that family of theories of perception which assert that we acquire knowledge of the physical world *via* perception (i.e., deny scepticism) and deny that our perceptually-acquired knowledge of the physical world is based, as a final court of appeal, on information about things other than physical objects perceived. In this paper I wish to develop, and examine some of the consequences of, such a theory. As it will turn out, some of these consequences are quite unexpected and perhaps even slightly bizarre, to the extent that some might regard the fact that such consequences follow from the theory as a *reductio* of the theory. But even if this should be so I submit that the effort is worth it, because for various reasons, all of which I provide in either an earlier paper¹ or this one (or both) I think any other direct realist theory must encounter even more serious difficulties. So if *my* theory fails I would contend that *any* direct realist theory will. Hence, depending on one's reaction to what transpires here, one might regard this paper as either a last-ditch vindication of, or a requiem for, direct realism.

Be that as it may, my procedure will be as follows: In Section I, I present my theory for the case of visual perception and briefly sketch some of the advantages it enjoys over some other versions of direct realism. In Section II, I expand my theory to cover the other perceptual modes. At the conclusion of Section II, I uncover what at first might seem to be a minor hitch in the scheme. But in Section III, where I briefly examine some recent controversies concerning the observation-theoretical distinction, it is seen that this hitch is really a special case of a more pervasive difficulty, and in

attempting to find a way around this difficulty I am led to adopt the unexpected and perhaps unpalatable consequences lately referred to.

I

Consider those visual situations described by "*x* looks *F* to *S*," where *S* is an observer, *x* a physical object that *S* sees, and *F* one of the purely visual properties, i.e., one of the colors. Clearly, *qua* being looked-to *F*, it is at least reasonable for *S* to believe, of *x*, that it is *F*. And it is important to notice that the reasonableness conferred upon this belief by *S*'s being looked-to *F* does not depend on the belief's being *true*: should *x* *not* be *F* it would still be *prima facie* reasonable (and in some situations even reasonable, all things considered) for *S* to believe that it is. It is important to keep this fact in mind, because it is one that seems to be overlooked by a number of accounts of perceptual reasons that have recently been making the rounds. These accounts vary in detail, but they all have this much in common: that knowledge-conferring perceptual reasons for believing that *p* are held to consist *solely* of a law-like relation obtaining between some existing state of the percipient and the state of affairs which makes '*p*' *true*.² But this fails to account for the point cited above: that there can be perceptual reasons for *false* beliefs as well as true ones. By way of reply, it may be argued that I have paid insufficient attention to the fact that the account of perceptual reasons under consideration is intended as an account of *knowledge-conferring* reasons only. And it is, of course, always possible to take the position that whatever it is about being-looked-to that makes some false perceptual beliefs reasonable is something that plays *no role whatsoever* in perceptual knowledge. But aside from the fact that such a position is implausible, it still leaves us with

¹ "Sensory Experience, Epistemic Evaluation and Perceptual Knowledge," *Philosophical Studies*, vol. 28 (1975), pp. 173-187.

² Alvin Goldman, "A Causal Theory of Knowing," *The Journal of Philosophy*, vol. 64 (1967), pp. 357-372; Peter Unger, "An Analysis of Factual Knowledge," *The Journal of Philosophy*, vol. 65 (1968), pp. 157-170; David Armstrong, *A Materialist Theory of the Mind* (New York, 1968) chapter 9, *Belief, Truth and Knowledge* (Cambridge, 1973) chapters 11-13; Fred Dretske, *Seeing and Knowing* (Chicago, 1969) chapters 3 & 4 and "Conclusive Reasons," *Australasian Journal of Philosophy*, vol. 49 (1971), pp. 1-22.

an unanswered question—namely, what is it about being-looked-to that makes some false perceptual beliefs reasonable? The foregoing “law-like relation” account of perceptual reasons does not provide us with an answer to this question.

But neither can the reasonableness conferred by x 's looking F to S upon S 's believing, of x , that it is F be explained solely in terms of inductive considerations. All knowledge of the purely visual properties ultimately stems from how things look, so reliance on induction alone could enable us at best to establish correlations among the looks of things, not between how they look and how they are. Clearly we must turn to something other than inductive rules. To this end it might be suggested that we simply adopt it as a fundamental epistemic principle that S is reasonable (or perhaps even fully justified) in believing, of something that looks F to him, that it is F , provided he has (at that time) no evidence to the contrary. But to adopt such a principle without explanation simply because it *seems* right is a rather arbitrary procedure. It would be better if we could analyze being-looked-to so as to *justify* adopting this principle.

The analysis I proposed in my earlier paper, and which I retain here, for accomplishing this end is perhaps best explained by contrast with an approach developed by Roderick Chisholm. In some of his writings on perception,³ Chisholm specifies, as one mark that some state of affairs is evidence for the truth of some proposition, that that state of affairs be describable without the use of terms that evaluate the worth of knowledge claims. Now what I wish to propose is that this condition cannot be fulfilled for perceptual knowledge. In particular, I wish to propose that the situation described by “ x looks F to S ” is evidence, for S , that something in his field of vision is F and, moreover, that it *cannot be described independently of its status as evidence, for S , that something in his field of vision is F* . So it follows that S is *prima facie* reasonable (or perhaps justified) in believing, of something that looks F to him, that it is F , because it is part of the *meaning* of being-looked-to that it provides S with such a *prima facie* reason (or justification).

But more needs to be said, because at this point it might be objected that to have evidence for the truth of some proposition, ‘ p ’, is at least in part to

believe some other proposition, ‘ q ’, whose truth is statistically or nomologically related to the truth of ‘ p ’. So it would seem that my analysis of being-looked-to ultimately reduces it to the holding of a *belief* (which, in turn, we should expect to be *prima facie* reasonable). If we are to avoid this conclusion, we must say that there is a way of having evidence—I will call it “perceptually-having” evidence—that is itself basic and cannot be understood in terms of the having of beliefs and/or the obtaining of observer-independent evidence-relations among states of affairs. The best we can do by way of explaining this way of having evidence is to point to examples. We can say, “Perceptually-having evidence that something in your field of vision is red is what happens when something looks red to you, or when it looks to you as if something in your field of vision is red.”

The foregoing theory thus explains, without the adoption of *ad hoc* epistemic principles, how being-looked-to confers reasonableness on visual beliefs. And it also enjoys some further advantages over many alternative theories of perception. Without presuming to provide much in the way of either detail or exhaustiveness (for more of both see my earlier paper) I might here briefly indicate what I take some of these advantages to be.

One kind of non-direct-realist theory of perception which has a long tradition is what I will call the “sensuous appearing” theory. According to such a theory, perception has a distinctively “sensuous” component, i.e., a component which is a non-conceptual, non-intentional⁴ state of the observer and which, *though* a component of perception, can occur when no physical object is being perceived (as in hallucinations). Furthermore, according to such a theory, our perceptually-acquired knowledge of the world is based, as a final court of appeal, upon non-inferential knowledge of this sensuous component. The difficulties encountered by theories of this sort are well-known and numerous, but at least this much is to be said in their favor: There *does* seem to be a component of perception—what, following standard practice, I will call “sensory experience”—which is not itself belief, but which provides reasons upon which perceptual beliefs are based. But it is precisely this fact about perception that so many direct realist theories seem unable to

³ “‘Appear,’ ‘Take,’ and ‘Evident,’” *The Journal of Philosophy*, vol. 53 (1956), pp. 729–739, and *Perceiving: A Philosophical Study* (Ithaca, 1957), pp. 34–35.

⁴ My use of “intentional” in this paper will reflect Roderick Chisholm’s definition of that term in Chapter 11 of *Perceiving*, except that I will use the term to refer to states and states of affairs (e.g., belief and so-and-so’s believing that p) rather than sentences.

accommodate. So David Armstrong⁵ and George Pitcher,⁶ e.g., have developed theories of perception which, allowing for minor embellishments and qualifications, view *x*'s looking *F* to *S* as *nothing but* *S*'s visually and non-inferentially acquiring the belief that something in his field of vision is *F*. And Moreland Perkins, while insisting that perception has an additional sensuous component, has denied that this component has any epistemological import.⁷

Now one advantage I claim for my theory of perception is that, on the face of it at least, it provides us with a more adequate account of sensory experience than the foregoing theories. This is done by identifying *S*'s perceptually-having evidence that something in his field of vision is *F* with *S*'s having a visual sensory experience of *F*. In this way we can identify sensory experience with something other than perceptual belief, but at the same time provide for it an epistemological role in the context of a direct realist theory of perception. But it should be noticed that, even though we have avoided identifying sensory experience with belief, we *have* identified it with an *attitude* toward the physical world on the part of the observer, an attitude which employs concepts and is intentional. It has been so widely assumed, at least by those who have been clear about it, that perception has a non-conceptual, non-intentional component and that it is *this* that is normally referred to by the terms "sense experience" and "sensory experience" (Armstrong and Pitcher have been the chief dissidents) that, simply on the theory that fifty million philosophers can't be wrong, one might be suspicious of any view to the contrary. Whether or not such a suspicion is justified is a matter that I will return to later in this paper.

II

Let us consider how the theory of sensory experience just proposed might be extended to modes of perception other than the visual. The main problem is to provide an account of what it is that enables us to distinguish among the various modes of sensory experience. A little reflection will show that a description of the physiology of each of the modes of perception is not what we are looking for. We cannot, e.g., say that what is distinctive

about visual experience is that the perceptually-having of evidence results from the action on light-sensitive organs of light reflected from physical objects, because in totally hallucinatory visual situations this just doesn't happen. And though it may well be true that visual experience is uniquely always a result of, or is even identical with, the appropriate stimulation of the appropriate portion of the central nervous system, we can certainly distinguish among the different sensory modes without an appeal to such physiological details. We can because we do. What I wish to know is how.

I see two possible ways of answering this question. I will call them the "state-difference" view (hereafter SD) and the "content-difference" view (hereafter CD). According to SD, the different sensory modes are *different intentional states*, just as belief, hope, and desire are different intentional states. So rather than there being *one* way of having evidence associated with perception—i.e., perceptually-having—there are several ways: visually-having, auditorily-having, and so forth. According to CD, on the other hand, there is only perceptually-having evidence, and the different sensory modes are distinguished in terms of which of the sensible *properties* and *relations* one perceptually-has evidence for the presence of.

A serious problem with SD is that, since different intentional states can have identical contents, SD commits us to the logical possibility of our having, e.g., visual *as well as* auditory experiences of sound, and auditory *as well as* olfactory experiences of smell. Since this hardly does seem logically possible, CD is the alternative I will adopt. In the balance of this section I propose to work it out in some detail. As I see it, there are two main difficulties to contend with. One has to do with the common sensible properties (and relations) such as shape. How can the different sensory modes be accounted for solely in terms of different properties when it is possible to have visual and tangible experiences of the same property? The second difficulty—namely, the seeming arbitrariness of specifying which properties correspond to which sensory modes—has been pointed out by H. P. Grice: "... to say that colour is a visual property would seem to amount to no more than saying that colour is a member of a group of properties the others of which

⁵ *Perception and the Physical World* (London, 1961) and *op. cit.*

⁶ *A Theory of Perception* (Princeton, 1971).

⁷ "Matter, Sensation and Understanding," *American Philosophical Quarterly*, vol. 8 (1971), pp. 1-12, and "Sentience," *The Journal of Philosophy*, vol. 68 (1971), pp. 329-337.

are . . . This leaves membership of the group as an apparently arbitrary matter."⁸

J. W. Roxbee Cox has recently proposed to deal with both problems in terms of a "key property" account of the differences among the senses.⁹ According to Cox, to each sensory mode there corresponds a *single* key property such that any *other* property that might be perceived *via* that mode must be perceived *through* perceiving the key property. So, e.g., color is the key property for vision, because whenever someone sees any other property he necessarily sees it *through* seeing colors. And though the issues here are complex, for our purposes I think it will suffice to say (and here I am going a bit beyond Cox's treatment of the matter) that the key property *through* which we feel shapes and spatial relations is "the presence or absence of kinaesthetic resistance."¹⁰ (And hereafter, when I talk about "tangible experiences of shape" and "kinaesthetic experiences of space" I will have in mind this mode of feeling shapes and spatial relations). Thus, the assigning of certain properties and relations to certain sensory modes is no longer an arbitrary matter; we assign to each mode the appropriate key property together with those properties and relations which can be perceived *through* perceiving it. And the difference between, say, visual and tangible-kinaesthetic experiences of the same property is explained as a difference in *how* (i.e., through what key property) this property is perceived.

But though I agree with Cox's approach as far as it goes, I don't think it frees us of the necessity of drawing a distinction that some philosophers are likely to regard as suspicious if not downright disreputable—namely, a distinction between visual and tangible shape, or at least between the concepts thereof (with similar distinctions for the other common sensible properties and relations). To see why this is so, consider the following:

- (1) *S* has, at (time) *t*, a visual experience of *G*.
- (2) *S* has, at *t*, a tangible experience of *G*.
- (3) *S* has, at *t*, a visual experience of color.
- (4) *S* has, at *t*, a tangible experience of kinaesthetic resistance

where *G* is any shape. If we adopt Cox's key property approach, and allow " $\square$ " to be the operator for logical or conceptual necessity, we get the following relations:

- (a) $(\forall S)(\forall t) \square [(1) \supset (3)] \ \& \ (\forall S)(\forall t) \sim \square [(1) \supset (4)]$
- (b) $(\forall S)(\forall t) \sim \square [(2) \supset (3)] \ \& \ (\forall S)(\forall t) \square [(2) \supset (4)]$

But if we adopt the theory of sensory experience I am proposing and CD, (1) and (2) are each to be rendered in terms of some substitution instance of propositional function *P*:

- P*. *S* perceptually-has, at *t*, evidence that something is present which is $\emptyset$

Now suppose, anticipating the results, we let "visually-*G*" and "tangibly-*G*" be those terms which, upon replacing the variable ' $\emptyset$ ', yield those substitution instances of *P* which are correct renderings of (1) and (2) respectively, and let those substitution instances be numbered '(1\')' and '(2\')'.

- (1') *S* perceptually-has, at *t*, evidence that something is present which is visually-*G*.
- (2') *S* perceptually-has, at *t*, evidence that something is present which is tangibly-*G*.

Now if we replace (1) and (2) with (1') and (2') in (a) and (b), we see that, for any given values of '*S*' and '*t*', (1') and (2') each entails a proposition not entailed by the other. And from this it seems appropriate to conclude that "visually-*G*" and "tangibly-*G*" are expressive of *different concepts*.

Now I think this line of reasoning is basically correct, but the final step calls for some defense. What we must keep in mind is that perceptually-having evidence, as I understand it, is intentional. Now suppose we let '*I*' be any verb which describes intentionality. Then the final step of the foregoing argument assumes that, if '*p*' and '*q*' are sentences which express the same proposition (or, if one prefers, are logically equivalent sentences, or logically equivalent propositions) then "*S I* that *p*" and "*S I* that *q*" must be so too. But there is a common use of intentional verbs for which we cannot correctly say that *S I* that *p* unless *S* himself would be willing to describe the content of his

⁸ "Some Remarks about the Senses," R. J. Butler (ed.) *Analytical Philosophy*, Series I (Oxford, 1962), p. 140.

⁹ "Distinguishing the Senses," *Mind*, vol. 79 (1970), pp. 530-550.

¹⁰ According to Cox, there are a number of different kinds of feeling (contact feeling, feelings of pervasive warmth or coolness, etc.) all of which are sub-species of a more general sensory mode which necessarily involves the perception of one's environment *through* perception of one's own body. But this may be correct (and I think it is) and still be the case that we can specify sub-key properties for all or some of these sub-species. And this is what I have tried to do for that mode of feeling by which we perceive shapes and spatial relations, with, I trust, at least a sufficiently close approximation to the truth to enable us to get on with it.

intentional state by use of the *assertion* or *inscription* 'p'. And on this use of intentional verbs the substitution principle just cited does not hold. But there is another common use of intentional verbs for which the story is different, as witness the fact that we can quite appropriately say, of someone who does not know a word of English, that he believes that it is raining outside. On this use of 'I', "*SI* that p" might read (depending on one's philosophical biases on these matters) something like, "*SI* I, of the proposition expressible (in such and such language) by 'p', that it is true." In this case, replacing 'p' by a sentence expressing the same proposition *does* retain the original meaning of the whole.

Now our seeing the shapes of things through perceiving colors and not through perceiving kinaesthetic resistance (i.e., (a) above) certainly does not depend on what particular assertions or inscriptions we would be willing to use to characterize this state of affairs. And the same also holds of our feeling the shapes of things through perceiving kinaesthetic resistance and not through perceiving colors ((b) above). So in saying that, for any given values of 'S' and 't', (1') and (2') each entails a proposition not entailed by the other, it is sufficient to interpret "perceptually-has evidence" in the latter of the two ways cited in the preceding paragraph. So we *can* legitimately draw the conclusion that the "visually-G" used for (1') and the "tangibly-G" used for (2') are expressive of different concepts. And these terms are certainly natural and appropriate ones to delineate this difference. And to *refer* to the concepts that are so expressed I will italicize, speaking of "the concept, *visual-G-ness*" and "the concept, *tangible-G-ness*."

The arguments and conclusion of the three preceding paragraphs have been developed specifically for the case of shape. But it is easy enough to see how they apply to any of the common sensible properties and relations, enabling us to distinguish, e.g., between the concepts, *visually-spatially related* and *kinaesthetically-spatially related*. In what follows I will confine my discussion largely to shapes, but it should be kept in mind that parallel conclusions hold for the other common sensible properties and relations as well.

But at this point an objection might be raised. It might be argued that if we are to take seriously the distinction between the concepts, *visual shape* and *tangible shape* we must allow the logical possibility

that different correlations could (sometimes or even always) hold between the visual and tangible sets of shapes than those that in fact do. We must allow, e.g., that the *same physical object* can be both visually cubical and tangibly spherical. And since this is not logically possible the distinction between the visual and tangible concepts cannot be sustained. But *are* such off-beat correlations logically impossible? Well, it may depend on how we understand it. Suppose the correlation were this: If someone were to place his hands around an object which standardly looks cubical, it would look just the way it looks to *us* when we place our hands in standard conditions, around an object which is cubical, but it would *feel* to him the way it feels to *us* when we place our hands, in standard conditions, around an object which is spherical. Judith Jarvis Thompson has constructed a rather ingenious example which poses a serious problem for the possibility of such correlations' holding.¹¹ Suppose we arrange three of the visually-cubical, tangibly-spherical objects so that the way they look is represented as in Figure 1. The way they feel would then presumably be represented as in Figure 2; in particular there

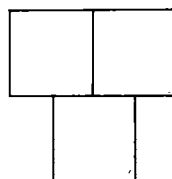


Fig. 1

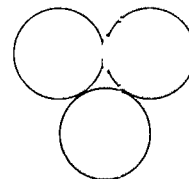


Fig. 2

would feel to be an empty space in the middle of the arrangement. Suppose now we hold a suitably small object over the middle and release it. Presumably it will feel as if it drops through to the floor, whereas it will look as if it terminates its fall on the top of the arrangement. And now the question arises: Just where *is* the object?

Now I take it what Thomson intends by this example is to show that, if we allow the envisioned off-beat correlations to hold between the visual and tangible sets of shapes, things would be so helter-skelter that the concept of relatively stable physical objects to which we can attribute properties would not apply. But I don't think Thomson has shown this. Rather, what she has shown is that, if we alter the existing correlations between visual and

¹¹ See her "Molyneux's Problem," presented at the 1974 Eastern Division Meetings of the American Philosophical Association and published in *The Journal of Philosophy*, vol. 71 (1974), pp. 637-650.

tangible shapes in the way envisioned, we must also alter the existing correlations between visual and kinaesthetic spatial relations—that, e.g., we must allow the occurrence of visual spatial contiguity to sometimes be accompanied by the occurrence of kinaesthetic spatial noncontiguity. But if we are to take seriously the distinction between the concepts, *visual spatial relation* and *kinaesthetic spatial relation* this sort of thing follows as a logical possibility anyway. Now certainly a world in which this occurred would be bizarre, but it is not at all obvious that it need be helter-skelter or unpredictable to the extent that we would have to abandon all talk of physical objects; it could well be possible to arrange things so that it would make perfectly good sense to say, of the *same pair of physical objects*, that they are sometimes visually-spatially related in one way and kinaesthetically-spatially related in a different way. But it is not really necessary to speculate on how this might go. We could instead understand the idea of visual cubicalness being correlated with tangible sphericity in a way different from that to which Thomson addresses herself.

Suppose that all and only those objects which have a spherical tangible surface have this surface completely submerged in a cubical field through which other objects (the tangible parts included) can pass without resistance, but which is completely opaque to light. So an object standardly looks cubical if and only if it standardly feels spherical, or in other words, visual cubicalness is constantly correlated with tangible sphericity. (We can even add that the tangible sphere is off-center from the visual cube, so that the visual and kinaesthetic spatial relations among these objects would also be different—though, of course, not in the way that we envisioned in the preceding paragraph.) If there is a reluctance to accept this on the grounds that the intangible, visually opaque field is too inconsequential to be considered a part of the object, we can always give this field causal properties sufficient to answer the objection. Once this is done we see that it is logically possible that different correlations could obtain between the visual and tangible sets of shapes than those that in fact do. So the fact that maintaining a distinction between the concepts, *visual shape* and *tangible shape* commits us to this

possibility cannot be used as an argument against the distinction.

But now let's consider another point. Our ordinary, common sense conceptual framework certainly seems to be one in which we have one concept of shape rather than two—let us call this common sense shape “proper shape.” We might now want to know how the concepts, *visual shape* and *tangible shape* (and specifically, *sphericity*, *cubicalness*, etc.) stand in relation to the concept, *proper shape*, (*sphericity*, *cubicalness*, etc.) Though I do not consider myself presently capable of giving anything like a complete answer to this question, the following scattered remarks may shed some light on the subject. First, we should resist the temptation to think of the concept, *proper shape* as being the result of some kind of “coalescing” of the visual and tangible concepts into one, or of the dominance or usurpation of the role of one by the other. For one thing, we can literally attribute proper shape to things to which we can attribute visual or tangible shape only in a Pickwickian sense at best—e.g., to a force field that is invisible and intangible and occupies a finite portion of space.¹² Furthermore, the proper shapes can in some cases be attributed to (possible) objects which exhibit a breakdown of the correlations that presently obtain between visual and tangible shapes. If, e.g., we allow our force field to be opaque to light but still intangible, it has visual shape but not tangible shape, but we can attribute to it proper shape, namely, the proper shape we would standardly see it to have. And in the imaginary world described one paragraph back we can attribute (indeed, I did attribute) proper shape to the visually-cubical and tangibly-spherical objects. We can say that they are objects consisting of a visible, intangible cubical field with a spherical hard core, where we mean “cubical” and “spherical” in the proper sense. But I think that we cannot attribute proper shapes to the visually-cubical and tangibly-spherical objects of Thomson's imaginary world discussed two and three paragraphs back.

These considerations, unsystematic and incomplete as they are, make it clear that the relation between the concept, *proper shape* and the concepts, *visual shape* and *tangible shape* is not a simple one. They also make it clear that we are now faced with

¹² In this and the following example I am relying on the principle that if there are *no* conditions under which we could perceptually-have evidence, with respect to some object, that it has some sensible property *P*, then *ipso facto* we could not perceptually-have such evidence under standard conditions; and if we could not perceptually-have such evidence under standard conditions, the object does not have sensible property *P*.

a problem. The distinction between the concepts, *visual shape* and *tangible shape* was originally introduced as a way of accounting for the difference between visual and tangible experiences of shape. But now it seems that this distinction avails us nothing, because since sensory experience is supposed to be conceptual, we should expect it to reflect the percipient's conceptual framework, and our conceptual framework seems to be one in which the distinction between the concepts, *visual shape* and *tangible shape* does not operate. Having climbed out on a limb, we seem to have sawed ourselves off. Faced with this difficulty we might try to show that, appearances to the contrary, the terms, "proper shape," "proper spatial relation," etc. *can* be appropriately defined in terms of, or reduced to, the terms for the corresponding visual and tangible-kinaesthetic properties and relations. But the effort would hardly be worth it, because, as I will show in the following section, the present difficulty points the way to a more general difficulty that affects my whole theory of sensory experience and is not amenable to such a solution.

III

Suppose a physicist and a scientifically illiterate layman are each shown a photograph of a Wilson cloud chamber experiment and asked to describe what they see. The layman replies that he sees a number of white lines arranged in such and such a way against a dark background. The physicist says that he sees the ionization of gas molecules marking the passage of alpha particles.

We may distinguish two different ways of characterizing this state of affairs. One characterization, representing a rather traditional point of view, is that whereas the layman is reporting what he directly and non-inferentially observes, the physicist is coupling the information gained from what he directly observes (this being the same as what the layman directly observes) with his knowledge of molecular theory to *infer* the occurrence of the

relevant molecular event. Now we can all agree that there must be *something* to the distinction we intuitively feel between, e.g., reporting the presence of a man because one sees the man and reporting the presence of a man because one sees his footprints (though I will argue later that we have to be very careful how we draw this distinction). But in the present kind of context a further distinction is often being exploited, one between not merely different kinds of *reports* but different *languages*. There is, according to this view, a pure observation language which employs categories drawn from and applicable to entities which are "in principle" observable, and a theoretical language which correspondingly applies to entities which are "in principle" unobservable. And it is only by using terms of the observation language that a *bona fide* observation report can be made.¹³

But there is a second view to be considered, one representing a more recent line of thinking, and according to this view no such observation-theoretical distinction can be drawn. *All* observations (or at least all descriptions of what we observe) are theory-laden, in the sense that they all presuppose commitment to a conceptual framework that functions for the observer in the same way that molecular theory functions for our physicist. In view of this (so it is held) the important distinction to be drawn is not between observation sentences and theoretical sentences, but between sentences whose truth values are "quickly-decidable" and those for which this is not so. And there is nothing sacrosanct about those sentences which are presently (for most of us) quickly-decidable. Should it be useful at some future time, we all *could* replace our ordinary conceptual scheme *in toto* with that of the most up-to-date physics, thereby reporting what we observe as, e.g., "an object with such and such a molecular structure in virtue of which it standardly reflects light of so and so wave length" rather than "an object of such and such color," or "a body of gaseous molecules with relatively high mean square velocity" rather than "hot air."¹⁴

¹³ Variations on this theme have been developed by, among others, N. R. Campbell, *Foundations of Science* (New York, 1957—originally published as *Physics, the Elements* in 1920) chapter 6; Ernest Nagel, *The Structure of Science* (New York, 1961) chapter 5; and Rudolph Carnap, "The Methodological Character of Theoretical Concepts," in *Minnesota Studies in the Philosophy of Science*, ed. by Herbert Feigl and Michael Scriven, vol. I (Minneapolis, 1956).

¹⁴ For views which embody many or all of the foregoing features see N. R. Hanson, *Patterns of Discovery* (Cambridge, 1958); T. S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1962); Paul Feyerabend, "An Attempt at a Realistic Interpretation of Experience," *Proceedings of Aristotelian Society*, vol. 58 (1957–58) pp. 143–170, "Explanation, Reduction and Empiricism," Herbert Feigl and Grover Maxwell (ed.) *Minnesota Studies in the Philosophy of Science*, vol. III (Minneapolis, 1962), "How to be a Good Empiricist," Bernard Baumrin (ed.) *Philosophy of Science: The Delaware Seminar*, vol. 2 (New York, 1963); and Wilfred Sellars, "The Language of Theories" (with commentary by Paul Feyerabend) Herbert Feigl and Grover Maxwell (ed.) *Current Issues in the Philosophy of Science* (New York, 1961).

All of this, of course, is familiar enough by now, but let's see how my theory stands with respect to these two competing views—call them the “traditional” and “non-traditional” views. On my theory there are certain properties and relations which play a special role in perception, and this role would seem to provide a basis for a distinction between observable and theoretical properties and relations. The observable properties and relations, namely, are those properties and relations for the presence of which one can perceptually-have evidence; the theoretical properties and relations are those for which this is not so. With this distinction as a basis, it is then an easy enough matter to construct an observation language of the sort envisioned by traditionalists. (Since our main concern here is with whether or not there are purely observational *descriptions*, we need not worry about drawing the observable-theoretical distinction for particular objects. The general terms can suffice for fulfilling the descriptive function of our observation language.) In the same spirit, we can specify a precise definition of an observation *report* such that such a report is formulable in the observation language. We can say that an observation report is a present tense report given by somebody of the content of his sensory experience.

Now I think it is arguable that, as far as non-logical vocabulary goes, and with possibly one important set of exceptions I will mention later, most of our everyday discourse about the everyday objects about us is formulable in a traditionally conceived observation language of the sort just suggested. But many of what we would normally call “observation reports” are *not* observation reports in the sense in which I have defined them. When, upon seeing an orange, one reports that there is an orange over yonder, one is reporting more than the content of his visual experience. The latter would be restricted to asserting the presence of something of a certain color and visual shape, and there is more to being an orange than just that. A comparison with the case of the physicist describing what he sees in the photograph of the Wilson cloud chamber experiment is instructive. In neither case is the percipient giving an observation report. Both reports go beyond the perceptual evidence, coupling this evidence with other evidence *via* induction, and (at least in the case of the physicist) whatever rules are relevant to the establishing of scientific theories. Furthermore, both reports are quickly-decideable for the respective observers. But there may be this much

difference: Whereas the report of the presence of the orange is formulable in the observation language, the scientist's report is not. (I say that there merely *may* be this difference because of the possible exception mentioned at the outset of this paragraph.) So, on my theory, there is this much to be said for the traditional view: that we can draw a distinction, in pretty much the way the traditional view claims, between observation and theoretical languages. But there is at least this much to be said for the non-traditional view: that as far as “quick-decideability” goes there is nothing sacrosanct about observation reports or even the observation language. Many of our everyday quickly-decideable reports about everyday objects are no more *observation* reports than are the equally (for some) quickly-decideable reports of physicists about objects of a more exotic nature. And many of the latter are quite clearly not even formulable in the observation language.

But it may be thought that there is implied in my theory a bigger concession to the non-traditional view than this. The observation-theoretical distinction, on my theory, stems from the fact that there are a certain limited number of things for which one can perceptually-have evidence. But perceptually-having evidence is a *conceptual* state of the observer, so why couldn't conceptual change issue in perceptually-having evidence for *different* things, and hence, to a different observation language? And yet, we know this wouldn't happen. Imagine, e.g., someone who, in the spirit of what has been suggested by some non-traditionalists, so internalizes the conceptual framework of modern physics that he responds to perceptual stimuli in the language thereof (and certainly such a thoroughgoing internalization is at least logically possible). He thinks of color in an object, e.g., not as a property spread continuously over the whole object, but as a certain molecular structure of the object whereby it reflects light of a certain wavelength. In consequence, he does not believe, as pre-scientific common sense does, that all parts of the object, no matter how small, are colored. He knows that the individual atoms of the object are not. But how would this person's sensory experience of, e.g., a red object, differ from ours? Certainly we know that it would differ not at all. He would perceptually-have evidence for the same thing that *we* perceptually-have evidence for—namely, a property that is spread *continuously* over the whole object. But, it will be said, that is because his eyes aren't sharp enough to see the molecular

structure. True enough, but not to the point. The point is, if perception is to be understood as essentially a *conceptual* ability, why doesn't a change in concepts result in a change in perception? Besides, the problem is not always one of things being too small to be observed. Suppose our scientist internalizes relativity theory and so comes to think and speak of length as a relation between a reference frame and an object rather than an intrinsic property of an object. How would his sensory experience differ from ours? Indeed, what on earth would a sensory experience of relational-length be like? But it probably isn't necessary to resort to hypothetical examples to make this point, because it should be clear from the discussion in the next to last paragraph of Section II that it is at least not *easy* to introduce such terms as "proper shape" and "proper spatial relation" into the observation language—and this is what I meant when I remarked earlier that our ordinary language may not entirely be an observation language. So *we* may well be in the position of verbally responding to our environment in the terms of a theoretical conceptual scheme. We respond in terms of the proper concepts, but on the sensory level the visual and tangible-kinaesthetic concepts remain. Indeed, to echo what was said a few lines back in connection with a hypothetical example, it is hard to imagine what a sensory experience of proper shape would be like.

The conclusion toward which these reflections seem to be pushing us is that, contrary to what I have been contending all along, sensory experience is non-conceptual. If so, we could (I would contend) retain a direct realist theory of perception only by divesting sensory experience of its epistemological role (and in this connection it is worth noting that Paul Feyerabend, one of the most outspoken advocates of the non-traditional view, seems to do just that.)¹⁵ Now I think there is a serious problem here—and of course it is as much of a problem for the Armstrong-Pitcher type of theory as for mine—and I am not entirely sure that it can be answered. But in what follows I wish to indicate the form I think a successful answer, if there is one, must take.

What I would say is this: We have an *innate conceptual framework* that manifests itself as a reaction

to perceptual stimuli, manifests itself in the way in which we perceptually-have evidence. This conceptual framework thus is precisely such as is necessary to account for the nature of our perceptual-havings of evidence, i.e., of our sensory experience. In particular it maintains a conceptual distinction between visual and tangible shapes, spatial relations and the like. But other conceptual frameworks can be superimposed upon this (let us call it) sensory one, and this is in two (related) ways. First, we can come to regard much of the evidence we acquire *via* the manifestation of the sensory framework as being misleading, not merely in the uninteresting sense that there are occasional perceptual errors but in the more systematic sense that, e.g. *nothing* is colored in the way in which our sensory experience would lead us to believe. Secondly, we can learn to make quick decisions about, and reports of, our environment employing conceptual frameworks other than the sensory framework—as I have suggested we may in fact do in everyday life. But the more primitive, sensory, framework remains relatively unchanged by such happenings and continues to manifest itself as reactions to perceptual stimuli. It also, of course, continues to provide us with the evidence upon which all our empirical knowledge about the world is ultimately based, even that knowledge which leads us to regard certain parts of the sensory framework as being systematically misleading. (And herein lies an answer to the objection that has often been raised against the non-traditionalists, that if scientific revolutions really involve the radical kind of conceptual changes they say, there is nothing left to the objectivity of science.)

Now I have no doubt that there are many who will find the foregoing paragraph a bit hard to swallow. The idea of an innate conceptual framework is certainly not something that is dear to the hearts of very many empiricists. Most would probably rather reject my theory of perception than accept that. But the trouble is, *any* direct realist theory of perception is going to run into the same or equally serious difficulties. *Any* direct realist theory, I would contend, must construe that component of perception by means of which perception puts us in contact with the world as an intentional and conceptual state.¹⁶ And if we are not to opt for the

¹⁵ See his "An Attempt at a Realistic Interpretation of Experience" *loc. cit.*, scattered portions of "Explanation, Reduction and Empiricism" *op. cit.*, and his commentary on Wilfred Sellars' "The Language of Theories," *op. cit.*

¹⁶ This isn't to say that there haven't been *attempts* to do otherwise. See G. J. Warnock, "Seeing," *Proceedings of Aristotelian Society*, vol. 55 (1954-55), pp. 201-218; Arthur Collins, "The Epistemological Status of the Concept of Perception," *The Philosophical Review*, vol. 76 (1967), pp. 436-459; and Fred Dretske, *op. cit.*, One of my reasons for thinking Dretske's theory won't

III. MIDDLE KNOWLEDGE AND THE PROBLEM OF EVIL

ROBERT MERRIHEW ADAMS

IF President Kennedy had not been shot, would he have bombed North Vietnam? God only knows. Or does He? Does even He know what Kennedy would have done?

There is a little known but interesting literature on the general issue exemplified by this question. In the 1580's a fierce controversy erupted between the Jesuits and the Dominicans about the relation between God's grace and human free will. The Jesuits held, among other things, that many human actions are free in the sense that their agents are not logically or causally determined to do them. ("Free" will always be used in this sense in the present essay.) How then does God maintain control over human history? Not by causally determining human actions, as the Dominicans seemed to believe,¹ but by causing circumstances in which He knew that we would *freely* act in accordance with His plans. This answer was developed with great ingenuity by Luis de Molina, and defended by other Jesuit theologians, notably by Francisco Suarez. Their theory includes the thesis that God knows with certainty what every possible free creature would freely do in every situation in which that creature could possibly find himself. Such knowledge was called "middle knowledge" by the Jesuits, because they thought it had a middle status between other kinds of knowledge—between God's knowledge of the merely possible and His knowledge of the actual; or between His knowledge of necessary truths, which all follow from the divine nature, and His knowledge of His own will and everything that is causally determined by His will.²

This paper is about two questions. The first is whether middle knowledge is possible, even for

God. I shall argue that it is not, on the ground that conditional propositions of the sort that are supposed to be known by middle knowledge cannot be true. I will examine (in section II) the attempts of Molina and Suarez to explain how God can have middle knowledge; and then (in section III) the account recently offered by Alvin Plantinga, who has reinvented the theory of middle knowledge. Two objections to my position will be discussed in section IV.

The idea of middle knowledge emerges in recent philosophical discussion chiefly because of its relevance to the second question that I shall discuss, which is whether God could have made free creatures who would always have freely done right. More precisely: Could God have brought it about that He had creatures who made free choices, but none of whom ever made wrong choices? The relevance of this question to the problem of evil is obvious and well known. If He could have, why didn't He? If He couldn't have, that's a good enough reason why He didn't. He could not have done it by causally determining the choices of creatures, for then their choices and acts would not have been free in the relevant sense. But it might seem that if God has middle knowledge, He could have secured creatures sinless but free by just creating those that He knew would not sin if allowed to act freely. In section V, therefore, we shall see what light the discussion of middle knowledge may shed on the question whether God could have arranged to have free creatures who were all sinless.

But first of all (in section I) I will try to explain why there seems to me to be a problem about the possibility of middle knowledge.

¹ An acutely argued Dominican contribution to the debate is Diego (Didacus) Alvarez, O.P., *De auxiliis divinae gratiae et humani arbitrii viribus, et libertate, ac legitima eius cum efficacia eorumdem auxiliorum concordia* (Rome, 1590); see especially the seventh disputation.

² I believe Molina originated the term "middle knowledge" (*scientia media*). I have given a very simplified account of his reasons for thinking it appropriate. See his *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione concordia* [hereafter abbreviated, *Concordia*], ed. by John Rabeneck (Oña and Madrid, 1953), qu. 14, art. 13, disp. 52, n. 9–10, and disp. 53, memb. 1, n. 6, and memb. 4, n. 4 (pp. 339f., 360, 394).

I

In the twenty-third chapter of the first book of Samuel it is written that after David had rescued the Jewish city of Keilah from the Philistines, and settled his men there, Saul made plans to besiege Keilah in order to capture David. When David heard of Saul's plans, he consulted God by means of an ephod, which apparently was an instrument of divination that yielded a yes-or-no answer to questions. David asked, "Will Saul come down, as thy servant has heard?" The Lord answered affirmatively. Then David asked, "Will the men of Keilah surrender me and my men into the hand of Saul?" And the Lord replied, "They will surrender you." Thereupon David evacuated his men from Keilah, and hid out in the hills, with the result that Saul did not have the opportunity to besiege him in Keilah, and the men of Keilah did not have occasion to betray him to Saul. (I Samuel 23:1-14, RSV)

This passage was a favorite proof text for the Jesuit theologians. They took it to prove that God knew the following two propositions to be true:

- (1) If David stayed in Keilah, Saul would besiege the city.
- (2) If David stayed in Keilah and Saul besieged the city, the men of Keilah would surrender David to Saul.

This is a case of middle knowledge; for it is assumed that all the actions mentioned in (1) and (2) would have been *free*, in the relevant sense, if they had occurred.

If we suppose that God is omniscient, we cannot consistently doubt that He had this middle knowledge unless we doubt that (1) and (2) were true. Therefore, as Suarez says, "the whole controversy comes back to this, that we should see whether those conditionals have a knowable determinate truth."³

But I do doubt that propositions (1) and (2) ever were, or ever will be, true. This is not because I am inclined to assert the truth of their opposites,

- (3) If David stayed in Keilah, Saul would *not* besiege the city.

- (4) If David stayed in Keilah and Saul besieged the city, the men of Keilah would *not* surrender David to Saul.

Suarez would say that (1) and (3), and (2) and (4), respectively, are pairs of contradictories, and therefore that one member of each pair must be true. He thus affirms what has been called the law of Conditional Excluded Middle. But this is a mistake. To obtain the contradictory of a conditional proposition is not enough to negate the consequent; one must negate the whole conditional, as was pointed out by Suarez's Dominican opponent, Diego Alvarez.⁴ It is true that in everyday speech we might deny (1) by asserting (3), as we may deny a proposition by asserting any belief we hold that is obviously enough inconsistent with it. But we might also deny both of them by asserting, "If David stayed in Keilah, Saul might or might not besiege the city." I believe the case of what Saul would or might have done if David stayed in Keilah provides a plausible counterexample to the proposed law of Conditional Excluded Middle; and philosophers have found even more convincing counterexamples.⁵

I do not understand what it would be for any of propositions (1)-(4) to be true, given that the actions in question would have been free, and that David did not stay in Keilah. I will explain my incomprehension.

First we must note that middle knowledge is not simple *foreknowledge*. The answers that David got from the ephod—"He will come down," and "They will surrender you"—are not understood by the theologians as categorical predictions. If they were categorical predictions, they would be false. Most philosophers (including Suarez but not Molina) have supposed that categorical predictions, even about contingent events, can be true by corresponding to the actual occurrence of the event that they predict. But propositions (1) and (2) are not true in this way. For there never was nor will be an actual besieging of Keilah by Saul, nor an actual betrayal of David to Saul by the men of Keilah, to which those propositions might correspond.⁶

Some other grounds that might be suggested for

³ Suarez, *De gratia*, prol. 2, c. 7, n. 1, in his *Opera omnia* (Paris, 1856-1878), vol. 7, p. 85. (All my page references to *De gratia* will be to this edition and volume.)

⁴ Alvarez, *op. cit.*, Bk. 2, disp. 7, n. 30 (p. 74). See Suarez, *De gratia*, prol. 2, c. 7, n. 24 (p. 95).

⁵ David Lewis, *Counterfactuals* (Oxford, 1973), p. 79f.; John H. Pollock, "Four Kinds of Conditionals," *American Philosophical Quarterly*, vol. 12 (1975), p. 53. The law of Conditional Excluded Middle was defended by Robert C. Stalnaker, in "A Theory of Conditionals," *American Philosophical Quarterly Monograph Series*, No. 2, *Studies in Logical Theory*, ed. by Nicholas Rescher (Oxford, 1968), p. 106f.

⁶ Suarez saw this point pretty clearly; see his "De scientia Dei futurorum contingentium" [hereafter abbreviated, DSDFC], Bk. 2, c. 5, n. 6 (*Opera omnia*, vol. 11, p. 357).

the truth of (1) and (2) are ruled out by the assumption that the actions of Saul and the men of Keilah are and would be free in the relevant sense. The suggestion that Saul's besieging Keilah follows by *logical* necessity from David's staying there is implausible in any case.⁷ It would be more plausible to suggest that Saul's besieging Keilah follows by *causal* necessity from David's staying there, together with a number of other features of the situation which in fact obtained. But both of these suggestions are inconsistent with the assumption that Saul's action would have been free.

Since necessitation is incompatible with the relevant sort of free will, we might seek non-necessitating grounds for the truth of (1) and (2) in the actual intentions, desires, and character of Saul and the Keilahites. It does appear from the Biblical narrative that Saul actually intended to besiege David in Keilah if he could. Perhaps proposition (1) is true by virtue of its correspondence with Saul's intention. One might also suppose that (2) was true by virtue of correspondence with the desires and character of the leading men of Keilah, if not their fully formed intentions. Maybe they were cowardly, untrustworthy, and ungrateful. And I take it that neither the Jesuits nor Plantinga would say that Saul's intentions, or the desires and character of the Keilahites, necessitated their actions or interfered in any way with their freedom of will.

But the basis thus offered for the truth of (1) and (2) is inadequate precisely because it is not necessitating. A free agent may act out of character, or change his intentions, or fail to act on them. Therefore the propositions which may be true by virtue of correspondence with the intentions, desires and character of Saul and the men of Keilah are not (1) and (2) but

- (5) If David stayed in Keilah, Saul would *probably* besiege the city.
- (6) If David stayed in Keilah and Saul besieged the city, the men of Keilah would *probably* surrender David to Saul.

(5) and (6) are enough for David to act on, if he is prudent; but they will not satisfy the partisans of middle knowledge. It is part of their theory that God knows infallibly what definitely would happen,

and not just what would probably happen or what free creatures would be likely to do.⁸

II

I trust that it is clear by this point that there is reason to doubt the possibility of middle knowledge. Those who believe it possible have some explaining to do.

In Molina's explanation the superiority of God's cognitive powers bears the heaviest burden. He holds "that the certainty of that middle knowledge comes from the depth and unlimited perfection of the divine intellect, by which [God] knows certainly what is in itself uncertain."⁹ This came to be known as the theory of "supercomprehension." According to it God's intellect so immensely surpasses, in its perfection, all created free wills, that it "supercomprehends" them—that is, it understands more about them than would be necessary merely to comprehend them.¹⁰ But as Suarez pointed out in rejecting the theory of supercomprehension, to comprehend something is already to understand about it everything that is there to be understood, and it is absurd to suppose that anyone, even God, could understand more than that.¹¹ Molina seems to want to say that what free creatures would do under various possible conditions is not there, objectively, to be known, but that God's mind is so perfect that He knows it anyway. But that is impossible. The problem to be solved is how the relevant subjunctive conditionals can be true, and nothing that may be said about the excellence of God's cognitive powers contributes anything to the solution of that problem.

Suarez offers what seems to me the least clearly unsatisfactory type of explanation for the alleged possibility of middle knowledge. He appeals, in effect, to a primitive understanding, which needs no analysis, of what is for the relevant subjunctive conditionals to be true. Consider a possible free creature, *c*, who may not ever exist, and a possible free action, *a*, which *c* may freely do or refrain from doing in a possible situation *s*. We are to consider *c*, not as actually existing, but as having "possible being" in the cause (God) that is able to produce *c*. So considered, according to Suarez, *c* has a property (a *habitus*, as Suarez puts it) which is

⁷ Suarez makes a similar point: DSDFC, Bk. 2, c. 5, n. 11 (p. 358).

⁸ See Suarez, DSDFC, Bk. 2, c. 1, n. 1-2, and c. 5, n. 9 (pp. 343f., 357f.).

⁹ Molina, *Concordia*, qu. 14, art. 13, disp. 53, memb. 3, n. 10 (p. 389f.).

¹⁰ *Ibid.*, qu. 14, art. 13, disp. 52, n. 11, 17 (pp. 341, 345).

¹¹ Suarez, DSDFC, Bk. 2, c. 7, n. 6 (p. 366f.).

either the property of being a possible agent who would in *s* freely do *a*, or the property of being a possible agent who would in *s* freely refrain from doing *a*. *c* has one of these properties, although there is nothing either internal or external to *c*, except the property itself, which would make or determine *c* to have one of these properties rather than the other. God has middle knowledge of what *c* would do in *s*, because God knows which of the two properties *c* has.¹²

Many philosophers would object to Suarez's ontology of merely possible entities, but perhaps one could develop a similar account of the relevant conditionals without such an ontology. God's *idea* of *c*, for example, is presumably an *existing* subject of properties. And one might ascribe to it, as a primitive property, the property of being an idea which, if it were satisfied by anything in *s*, would be satisfied by an agent that freely did *a* in *s*. This would have the disadvantage, however, of implying that whether *c* would do *a* in *s* depends, not on a property of *c*, but on a property of God's idea of *c*. That consequence might seem to compromise *c*'s freedom of will.

My principal objection to Suarez's defense of the possibility of middle knowledge is not based on ontological considerations, however. I do not think I have any conception, primitive or otherwise, of the sort of *habitus* or property that Suarez ascribes to possible agents with respect to their acts under possible conditions. Nor do I think that I have any other primitive understanding of what it would be for the relevant subjunctive conditionals to be true. My reason for saying that Suarez's defense is of the least clearly unsatisfactory type is that it is very difficult to refute someone who claims to have a primitive understanding which I seem not to have.

III

In his several published discussions of the "free will defense" to the problem of evil, Alvin Plantinga has assumed, in effect, that God can have middle knowledge; and in the most recent of these dis-

cussions he has defended this assumption.¹³ Following Robert Stalnaker and David Lewis, Plantinga adopts what he calls "the possible worlds explanation of counterfactuals."¹⁴ For proposition (1) to be true, according to Plantinga's theory, is for the following to be the case:

- (7) The actual world is more similar to some possible world in which David stays in Keilah and Saul besieges the city than to any possible world in which David stays in Keilah and Saul does not besiege the city.

There are two important reasons for denying that this analysis establishes the possibility of middle knowledge.

(A) To the extent that it is plausible, the possible worlds explanation does not really give us a new solution to our problem about the truth of the crucial conditionals. It merely offers us a new and up-to-date form for the expression of attempted solutions that we may already have considered and rejected. (In fairness it should be said that Plantinga does not claim otherwise.) Two points must be made here.

(i) If the explanation is to be plausible, the kinds of similarity among possible worlds that are allowed to be relevant to the truth and falsity of counterfactual conditionals must mirror the considerations that would in any case determine our judgment of their truth and falsity. Some similarities cannot plausibly be allowed any relevance at all. Among the possible worlds in which David stays in Keilah, for example, I suspect the most similar to the actual world is one in which Saul does not besiege Keilah, and in which the subsequent history of David, Saul, and of Israel and Judah goes very much as it did in the actual world. Perhaps in such a world Saul has a slightly different character, or acts out of character in a way that he does not in the actual world; but I doubt that that is as great a dissimilarity as the dissimilarity between a world in which there is a siege of Keilah by Saul (and perhaps a killing of David by Saul) and a world in which there is not. I certainly would not conclude, however, that therefore Saul would not have

¹² I believe this is what Suarez's views come to, as they are found in *De gratia*, prol. 2, c. 7, n. 21, 24, 25 (pp. 94-96).

¹³ The assumption passed unquestioned in Alvin Plantinga's *God and Other Minds* (Ithaca, 1967), ch. 6. In his *The Nature of Necessity* (Oxford, 1974), ch. 9, and less fully in "Which Worlds Could God Have Created?" *The Journal of Philosophy*, vol. 70 (1973), pp. 539-552, it is defended. At the same time Plantinga has attempted (successfully, I think) to free a part of his larger argument from dependence on the assumption (*The Nature of Necessity*, *op. cit.*, pp. 182-184). Plantinga has not used the term "middle knowledge," although it seems to me very apt for the expression of his views.

¹⁴ *The Nature of Necessity*, *op. cit.*, p. 178. See also Stalnaker, *op. cit.*, and Lewis, *op. cit.* In the present paper I shall disregard complications having to do with conditionals whose antecedents are impossible, as all the conditionals that will concern us have possible antecedents.

besieged Keilah if David had stayed in the city.¹⁵ That a world in which Saul besieges Keilah is in that respect unlike the actual world, is irrelevant to the question what Saul would have done if David stayed in Keilah. Some similarities between the actual world and other possible worlds are relevant to that question—for example, similarities and dissimilarities in causal laws and in people's characters. But we have already considered and rejected the idea of founding the truth of our crucial conditionals on causal laws or on people's characters.

(ii) Even the similarities that are allowed to be relevant to the truth of counterfactuals must not be given more decisiveness than we would otherwise accord to the considerations that they mirror. A world in which David stays in Keilah and Saul besieges the city is perhaps more similar to the actual world in respect of Saul's character than a world in which David stays in Keilah and Saul does not besiege the city. But we had better not conclude that therefore the former is more similar to the actual world than the latter for purposes of the possible worlds explanation, if we mean to adhere to the explanation. For this conclusion would give us more reason to reject the analysis in terms of similarity of possible worlds than to abandon our previous judgment that Saul *might* have acted out of character and so would only *probably*, not definitely, have laid siege to Keilah if David had stayed in the city. The issue here is a general one, and important. We have a well entrenched belief that under many counterfactual conditions many a person *might* have acted out of character, although he probably would not have. If the possible worlds explanation is to be plausible, it must not give such decisiveness to similarities of character and behavior as to be inconsistent with this belief.

(B) On the possible worlds theory, moreover, the truth of the crucial conditionals cannot be settled soon enough to be of use to God. The chief importance of middle knowledge, for Plantinga as well as Molina and Suarez, is that God is supposed to be guided by it in making decisions about the creation and providential governance of the world. And as Molina and Suarez insist, if God is to make such use of it, His middle knowledge must be prior,

if not temporally, at least in the order of explanation (*prius ratione*, as Suarez puts it), to His decisions about what creatures to create.¹⁶ For similar reasons the truth of the conditional propositions which are the object of middle knowledge must not depend on God's creative decisions. Ignoring angels (fallen or unfallen) for the sake of argument, let us suppose that Adam and Eve were the first free creatures that God made. We are to think of God as choosing from among many alternatives; among them were creating Adam and Eve, creating other free creatures instead of them, and making no free creatures at all. According to the theory of middle knowledge, God's decisions to make some free creatures, and Adam and Eve in particular, are to be explained in part by the truth of

- (8) If God created Adam and Eve, there would be more moral good than moral evil in the history of the world.¹⁷

This explanation would be viciously circular if the truth of (8) were later in the order of explanation than the decisions it is supposed to help explain.

Here we are dealing with a type of subjunctive conditionals that we may call *deliberative conditionals*. They ought not, in strictness, to be called *counterfactual*. For in asserting one of them one does not commit oneself to the falsity of its antecedent. That is because a deliberative conditional is asserted (or entertained) in a context of deliberation about whether to (try to) make its antecedent true or false. In asserting such a conditional one commits oneself rather to the view that its truth is independent of the truth or falsity of its antecedent.

There is a problem, which so far as I know has not been discussed in the literature, about applying to deliberative conditionals, as Plantinga does, the possible worlds explanation of counterfactuals.¹⁸ Consider a deliberative conditional,

- (9) If I did *x*, *y* would happen.

Is (9) true? According to the possible worlds explanation, that depends on whether the actual world is more similar to some world in which I do *x* and *y* happens than to any world in which I do *x* and *y* does not happen. That in turn seems to

¹⁵ Similar problems are discussed by Plantinga, *The Nature of Necessity*, *op. cit.*, pp. 174–179, and Lewis, *Counterfactuals*, *op. cit.*, pp. 72–77, 91–95.

¹⁶ See especially Suarez, DSDFC, Bk. 2, c. 4, n. 6, and c. 6, n. 3, 6 (pp. 355, 361, 363).

¹⁷ I have simplified here, particularly in the antecedent. God is supposed to have known that there would be more moral good than moral evil in the world if He executed a long series of actions, beginning with the creation of Adam and Eve. Many of these actions would be occasioned in part by responses He supposedly knew creatures would freely make to earlier actions in the series.

¹⁸ Stalnaker would apply it to deliberative conditionals too. Lewis might not; see his *Counterfactuals*, *op. cit.*, p. 4.

depend on which world is the actual world. And which world is the actual world? That depends in part on whether I do x . Thus the truth of (9) seems to depend on the truth or falsity of its antecedent. Similarly the truth of (8) will depend on whether God creates Adam and Eve.

I think it may be possible for a possible worlds theory of deliberative conditionals to overcome this difficulty in general, but not in such a way as to rescue the doctrine of middle knowledge. There is, I presume, a large class, K , of possible worlds that are more similar to some world in which I do x and y happens than to any world in which I do x and y does not happen. According to the possible worlds theory the truth of (9) depends on the actual world being some member of K , but not on *which* member of K it is. In asserting (9) in the context of deliberation I commit myself, in effect, to the view that the actual world is a member of K and that its membership in K does not depend on which I choose of the alternatives among which I am deliberating. This view may well be correct—if, for instance, x and y are linked by a strict causal law.

Similarly there is a class, K^* , of possible worlds that are more similar to some world in which God creates Adam and Eve and there is more moral good than moral evil in the history of the world than to any world in which God creates Adam and Eve and there is not more moral good than moral evil in the history of the world. The truth of (8) depends on the actual world being some member of K^* , according to the possible worlds theory. But how can the actual world's membership in K^* have been settled earlier in the order of explanation than God's decision whether to create Adam and Eve, or some other free creatures, or none? Here we face all the old difficulties about middle knowledge, and the possible worlds theory does nothing to help us answer this question. At most it explains why (8) is true, *given* that some member of K^* is actual.

Furthermore there is reason to believe that the actual world's membership in K^* cannot have been settled earlier in the order of explanation than God's decision. Let us say that one of God's alternatives is *represented* in K^* if and only if there is some world in K^* in which He chooses that alternative. If any of the alternatives among which God was choosing is not represented in K^* , then the actual world's membership in K^* depends on His rejecting that alternative, and therefore cannot be prior in the order of explanation to His decision.

But I think at least one of God's alternatives is indeed unrepresented in K^* . For one alternative was to make no free creatures at all, and I do not see how a world in which there are no free creatures at all could be a member of K^* . Since it is free actions that are morally good and morally evil,¹⁹ no possible world, w , will be a member of K^* unless there is some feature of w by virtue of which a difference in the free actions of free creatures in some worlds u and v would be a reason for counting u as more similar than v to w (in relevant respects). And any such feature of w must surely involve the existence in w of free creatures. If there are no free creatures at all in w , what would make w more like a world in which most free creaturely decisions are good ones than like a world in which most free creaturely decisions are bad ones? I conclude that the actual world's membership in K^* cannot be earlier in the order of explanation than God's decision to make some free creatures. Therefore the truth of (8), on the possible world's analysis, cannot be prior in the order of explanation to that decision.

Perhaps it will be objected to me that the partisans of middle knowledge need not claim that the truth of (8) precedes God's creative choices in the order of explanation. It is enough for their explanations if God *believed* (8) prior to making the choices. My reply is that if God acted on a belief in (8) before it was settled that (8) is true, then the fact (if it is a fact) that there is more moral good than moral evil in the history of the world is due to God's good luck rather than His wisdom—whereas the chief motivation of the theological theory of middle knowledge has been the desire to maintain that such happy results of God's dealings with created freedom are due to His wisdom, and that He had no need at all of luck.

IV

Of the philosophical objections that may be raised against my critique of the theory of middle knowledge, two seem to me the most important.

(A) I have relied on the claim that in the circumstances assumed in our example about David and Saul at Keilah, what is true by virtue of Saul's intentions and character is not

- (1) If David stayed in Keilah, Saul would besiege the city,

¹⁹ Plantinga insists on this point (*The Nature of Necessity*, *op. cit.*, p. 166f.).

but

- (5) If David stayed in Keilah, Saul would *probably* besiege the city.

Suarez has an interesting objection to this claim. He argues, in effect, that (5) can only mean that (1) is probably true, and that in accepting (5) one commits oneself, albeit with some trepidation, to the truth of (1).²⁰ Certainly it would be pragmatically inconsistent to assert that (1) is probably true and deny (as I do) that there is any way in which (1) can be true.

In proposing (5) as an alternative to (1), however, I do not understand it as a claim that (1), or any other proposition, *is* probable. It is rather a claim that

- (10) Saul will besiege Keilah

would be probable, given facts that would (definitely, not just probably) obtain if David stayed in Keilah. While "probably" is an epistemological term, moreover, it is used in (5) primarily to characterize dispositions or tendencies toward the truth of (10) that there would be if David stayed in Keilah. (5) does not imply that anyone would *know* the facts that would probabilify (10), but only that they would *obtain*, if David stayed in Keilah.

This view is consistent with treatment that (5) might receive under either of the two major types of theory of counterfactuals distinguished by Lewis. According to a *metalinguistic* theory, as Lewis puts it, "a counterfactual is true, or assertable, if and only if its antecedent, together with suitable further premises, implies its consequent."²¹ Holding a theory of this type, we might say that (5) is true if and only if (10) would be probable on total evidence constituted by the antecedent of (5), together with suitable further premises. The suitable further premises in this case would be partly about Saul's intentions and character. Lewis has proposed for the *possible worlds* theory an essentially similar treatment of counterfactuals that involve probability in the way that (5) does.²²

(B) Probably the most serious grounds for misgivings about my argument may be found in cases in which we seem to have confidence in what looks like a piece of middle knowledge. Suarez appeals

to such confidence on the part of ordinary speakers,²³ and Plantinga endeavors to provide us with convincing examples of it.

In one of Plantinga's fictitious examples Curley Smith, a mayor of Boston, has accepted a bribe of \$35,000 to drop his opposition to a proposed freeway route. In this case is the following true?

- (11) Smith would still have accepted a bribe to drop his opposition, if the bribe had been \$36,000.

Plantinga thinks "the answer seems fairly clear: indeed [Smith] would have" accepted the larger bribe;²⁴ and I agree.

But what makes (11) true? Let us note that it belongs to the class of subjunctive conditionals with antecedents assumed to be false and consequents assumed to be true, which have been called *semi-factuals*. What makes (11) true, I think, is that its consequent is true and the truth of its antecedent would not have prevented, or made less likely, the event that makes the consequent true. My view here is in accord with Nelson Goodman's claim that "in practice full counterfactuals affirm, while semifactuals deny, that a certain connection obtains between antecedent and consequent."²⁵ My account of what makes (11) true does not suggest a way in which (1) or (2) could be true, since they do not have true consequents to help make them true.

Furthermore, if my account is right, it was presumably not settled that (11) is true before (in the order of explanation) it was settled that Smith was going to be offered, and accept, \$35,000, since his actual acceptance is part of what makes (11) true. I see no reason, therefore, to suppose that God could have known of the truth of (11) early enough in the order of explanation to make use of it as He is supposed to make use of middle knowledge.

Another type of case, not presented by Plantinga, perplexes me more. There does not normally seem to be any uncertainty at all about what a butcher, for example, would have done if I had asked him to sell me a pound of ground beef, although we suppose he would have had free will in the matter.

²⁰ Suarez, DSDFC, Bk. 2, c. 5, n. 9 (p. 357f.). I am simplifying here, but I think not in such a way as to make this argument less plausible.

²¹ Lewis, *Counterfactuals*, *op. cit.*, p. 65.

²² David Lewis, "Counterfactuals and Comparative Probability," *Journal of Philosophical Logic*, vol. 2 (1973), p. 437f.

²³ Suarez, DSDFC, Bk. 2, c. 5, n. 8 (p. 357).

²⁴ Plantinga, *The Nature of Necessity*, *op. cit.*, p. 177.

²⁵ Nelson Goodman, *Fact, Fiction, and Forecast* (London, 1954), p. 15.

We say he would certainly have sold me the meat, if he had it to sell. What makes us regard it as certain? Chiefly his character, habits, desires, and intentions, and the absence of countervailing dispositions. (He would have had no motive to refuse me.)

There are three alternative views one might take of this case. One might say that if I had asked the butcher to sell me the meat, (i) he would only probably have sold it to me, though we normally ignore the minute but real chance there would have been that he would refuse; or (ii) he would certainly have sold me the meat, because he would have been causally determined to do so by his character and dispositions; or (iii) his character and dispositions would not have causally determined his action, but they render it absolutely certain that he would have complied with my request.

I have rested an important part of my argument on the assumption that what a person's character and dispositions do not causally determine, they do not render absolutely certain. Alternative (iii) is inconsistent with this assumption. It still seems to me, however, that my assumption is sound and alternative (iii) is more implausible than (i) or (ii)—although I must admit that I am not altogether content with either of them. For what is the nature of the rendering certain in alternative (iii), if it is not causal determination? On some views—Humean views—of the nature of probability and causality, alternative (iii) is plainly impossible; and I do not know of any theory that would render it intelligible.

V

Could God have arranged to have creatures who would perform free actions but only right ones? Let us consider the question first on the assumption that God has middle knowledge. In that case, we might think, He could have obtained sinless free creatures simply by making only those that He knew would always freely do right in those situations in which He would permit them to act freely.²⁶ Plantinga's response to this argument, a response which he develops with much greater elegance than I have space to reproduce here, is that God could not do this unless there are some possible free creatures who would in fact behave so well, and

that perhaps none would. Plantinga proposes the hypothesis that all possible free creatures (or their essences) have *trans-world depravity*. Roughly speaking, a possible free creature (or its essence) has trans-world depravity, in Plantinga's sense, if and only if that creature *would* do some wrong if God created it and permitted it to act freely, no matter what else God did. If the hypothesis of universal trans-world depravity is true, God must have known it is true, if He had middle knowledge, and must therefore have known that some evil was the inescapable price of created freedom.

Plantinga does not claim that the hypothesis is true, or even that it is plausible.²⁷ He argues only that it is logically possible, because he is using it to defend the view that it is logically possible that both God and evil exist. I do not doubt that the latter is logically possible; but religious thought must seek an account of the relations between God and evil that is credible, as well as logically possible.

It is worth asking, therefore, whether the hypothesis of universal trans-world depravity is plausible, on the assumptions about truth of conditionals that Plantinga shares with the Jesuit theologians. I think Molina and Suarez would deny that any possible free creature (or any free creature's essence) has trans-world depravity; and they could support their denial with persuasive arguments. Suarez holds that "it is alien to the common doctrine . . . and to the divine perfection and omnipotence, and is therefore of itself incredible enough, to say that God cannot predetermine [*praedefinire*] an honorable free act, in particular and with all [its] circumstances, by His absolute and effective will, the freedom of the created will still being preserved."²⁸ God uses his middle knowledge to make such predeterminations effective, choosing conditions and helps of grace that He knows will elicit a favorable response, and avoiding those under which He knows that the creature would not act according to the divine purpose. This presupposes, of course, that for every possible honorable free act of every possible free creature, in any possible outward circumstances, there are some incentives or helps of grace that God could supply, to which the creature would respond favorably though he could have responded unfavorably. But this is a very plausible presupposition if we assume, as Suarez does, that the theory of middle knowledge

²⁶ This argument is crisply stated by Nelson Pike, "Plantinga on the Free Will Defense: A Reply," *The Journal of Philosophy*, vol. 63 (1966), p. 93f. "Will" replaces "would" in Pike's formulation, but it is clearly middle knowledge that is involved.

²⁷ Plantinga, *The Nature of Necessity* (*op. cit.*), compare p. 165 with p. 189.

²⁸ Suarez, DSDFC, Bk. 2, c. 4, n. 4 (p. 354).

is correct, and that there is an infinite variety of natural and supernatural ways in which God can work on us inwardly, assisting our reasoning, affecting our feelings and perhaps our beliefs and desires, without causally determining our response.²⁹

And if it is plausible to suppose that for every possible *particular* occasion of action there are possible divine operations that would elicit a favorable free response, is it not also plausible to suppose that for many possible free creatures, and even for whole worlds full of them, there are possible series of divine operations to which those creatures would respond by *always* freely doing right, never doing wrong? Molina held that both Jesus and Mary were preserved from all sin throughout their whole lives by God supplying them with gifts and aids that He knew would always elicit a favorable free response from them.³⁰ Presumably He could have done the same for others.

If the hypothesis of universal trans-world depravity is implausible, it might seem that I offer theodicy a better alternative. I deny the possibility of middle knowledge, because I deny that the relevant subjunctive conditionals are true. In particular, I deny that the following is true:

- (12) If God had acted differently in certain ways, He would have had creatures who made free choices, but none of whom ever made wrong choices.

In other words, I deny that God could have made

free creatures who *would* always have freely done right. The supposition that He could have done so is burdened with all the difficulties about truth of conditionals that afflict the theory of middle knowledge. Since (12) is not true a reproach against God cannot rightly be based on its truth. And God cannot know that (12) is true, and cannot rightly be blamed for not using such knowledge.

My views about the truth of conditionals, however, do not tend to show that the following could not be true:

- (13) If God had acted differently in certain ways, He would *probably* have had creatures who made free choices, but none of whom ever made wrong choices.
 (14) If God had acted differently in certain ways, He would *probably* have had better behaved free creatures, on the whole, than He actually has.

In fact (13) seems to me rather implausible. Without middle knowledge God must take real risks if He makes free creatures (thousands, millions, or trillions of risks, if each free creature makes thousands of morally significant free choices). No matter how shrewdly God acted in running so many risks, His winning on *every* risk would not be antecedently probable. But I think (14) is very plausible. These judgments suggest that the necessity of permitting some evil in order to have free will in creatures may play a part in a theodicy put cannot bear the whole weight of it, even if the possibility of middle knowledge is rejected.³¹

University of California, Los Angeles

Received January 14, 1976

²⁹ Cf. *ibid.*, Bk. 2, c. 4, n. 5 (p. 355).

³⁰ Molina, *Concordia*, qu. 14, art. 13, disp. 53, memb. 4, n. 15-24 (pp. 399-405).

³¹ I am indebted to several, including David Kaplan, and especially David Lewis and Alvin Plantinga, for discussion and for comments on an earlier version of this paper, which was read to an American Philosophical Association symposium. An abstract of the earlier version, "Middle Knowledge," appeared in *The Journal of Philosophy*, vol. 70 (1973), pp. 552-552. Work on the present version was supported by the U.S. National Endowment for the Humanities.

IV. SKEPTICISM, FOUNDATIONALISM, AND PRAGMATISM

JOSEPH MARGOLIS

SKEPTICISM, G. E. Moore remarks, in a well-known essay,¹ "consists in holding, with regard to one particular *sort* of thing, that no human being ever knows with complete certainty anything whatever of that sort"; different forms of skepticism, Moore supposed, concern only the "sorts of things" about which the claim is supposed to hold. Moore took Russell to be the arch-advocate of skepticism so construed. He tried to show that Russell's arguments were inconclusive—and many of them clearly are; but Moore also insisted "that I have often known *with complete certainty* things of all these . . . sorts" (there are four distinct sorts of "things" that Moore considered and that Russell thought impossible to know with certainty) *and* he supposed that, in demonstrating the inconclusiveness of Russell's arguments, he was somehow showing that he *did* have knowledge with complete certainty of instances of the kinds of things in question. But there are enormous difficulties with Moore's position.

For one thing, Moore never explained what he meant by "complete certainty." It looks as if complete certainty must be either a distinct psychological state, which then would bear only a contingent connection with being in a "state" of knowledge, or else a distinct psychological state that qualifies, on some account, as meeting the conditions of knowledge plus further epistemic conditions regarding possible challenges to the justifiability of one's *belief* (belief that is thought to count as knowledge).² Secondly, Moore wished to make it very clear that he rejected "the common opinion that doubt is essential to skepticism," which

he thought arose "from the mistaken opinion that if a man sincerely believes that a thing is doubtful he must doubt it." Hence, Moore supposed that the *doubtfulness* of any of the sorts of "things" that the skeptic was making his claim about, had nothing to do with any psychological states at all.³ Nevertheless, he never managed to explain what was meant by "doubtfulness" or what its connection with "complete certainty" might be. But he obviously viewed doubtfulness in terms of the availability of epistemically appropriate challenges to the *truth* of what was claimed. Thirdly, Moore does not seem to have sufficiently appreciated the fact that *nothing* that he had to say confirmed that we knew anything with complete certainty. For example, his conclusion against some of Russell's detailed arguments is merely that it is logically possible that knowledge based on "analogical or inductive argument" might be (for all anyone could tell) "certain knowledge"; and that (as an instance) "it seems to me *more* certain that I *do* know that this is a pencil and that you are conscious than that any single one of [Russell's] questioned assumptions is true, let alone all four." But of course, the defeat of skepticism does not follow from the logical possibility that it might be false; and the fact that certain epistemic claims are *more certain* than Russell's skeptical assumptions does not entail that there are *any* truth claims that can be known with *complete certainty*. Moore also says, incidentally, that "I think I do know *immediately* things about myself." He also seems to hold that "memory is immediate knowledge" and seems to take it that

¹ G. E. Moore, "Four Forms of Skepticism" in *Philosophical Papers* (London, 1959). Further references to Moore are to this paper.

² Cf. Joseph Margolis, *Knowledge and Existence* (New York, 1973), Ch. 1.

³ There is, in fact, a serious difficulty in Moore's account of *what* these doubts are about. Moore rejects both facts and propositions and never says what the relevant "that"-clauses designate. But it is clear in context that he simply failed to notice an important equivocation. Thus he says: "Consider the form of words 'That the sun is larger than the moon is a fact which nobody knows for certain.' Here the proposition 'That the sun is larger than the moon is a fact' is logically equivalent to the proposition 'The sun is larger than the moon.' But the proposition 'Nobody knows for certain that the sun is larger than the moon' is logically equivalent to the proposition 'It is possible that the sun is not larger than the moon.'" But the second might well be read as "Though it is a fact (an accepted fact) that the sun is larger than the moon, nobody knows that a fact for certain (that is, no one knows that the putative fact, accepted as fact, is a fact for certain)." There is a similar quibble that arises with Moore's view of true propositions. But the upshot is that the skeptical position is here unaffected by Moore's maneuvers.

immediate knowledge is known with complete certainty. But, for one thing, he never demonstrates that he actually *has* immediate knowledge; he even concedes that he "thinks" that he does; and he never explains the sense in which memory (even of inner mental status) could be said to be a form of immediate knowledge, in the sense of providing knowledge with complete certainty. The upshot is that Moore's position is utterly inconclusive.

But there is more to the issue, because it may well be that the *kind* of skepticism that Moore was attacking is entirely benign—has no untoward consequences for the prospects of knowledge. In fact, Russell's claims are themselves entirely unaffected by certain detailed weaknesses of his own maneuvers that Moore properly uncovers in exploring the prospects of certainty. For instance, Russell seems to have thought that one could remember and yet be mistaken in what one remembered; and that, because one could be led into error in instances of certain kinds of cognitive claims, one could not be completely certain in other instances.⁴ But beyond this, it is useful to note that Keith Lehrer, for one, has actually embraced the kind of skepticism that Moore was attacking. As Lehrer says, "Not only do I doubt that anyone knows for certain that any [contingent] statement[s are] true. I also doubt that the lack of such knowledge is a serious epistemic loss."⁵ Indeed, Moore's position is a curious one since he actually allows the skeptic to admit that we have knowledge of the "sorts of things" about which we cannot, on the thesis, have knowledge "with complete certainty" (for instance, knowledge of one's own mental states; memory; knowledge of another's mental states; knowledge of external objects and events). But if Moore concedes the distinction, then he cannot take his claims against Russell to confirm that we (or he) know(s) anything with complete certainty. For his part, Lehrer embraces the thesis that "the skeptic is correct . . . in affirming the chance of error is always genuine."⁶ This is to concede Russell's

claim (which harkens back to Descartes' claim) that error is at least always logically possible. Lehrer's claim, however, goes beyond this minimal concession to argue a stronger *empirical* thesis, roughly, that the world is so disposed that there is a determinate propensity toward error. Hence, Lehrer assigns to the skeptic a more powerful claim than does Moore, in fact a claim that is more familiarly skeptical than is Moore's: "To sustain skepticism, a skeptic must go on to argue that if there is some chance that *S* is incorrect in his belief that *p*, then *S* does not know that *p*."⁷

The plot thickens considerably, therefore. For one thing, there is every reason to believe that Russell was *not* a skeptic in the clear sense that Lehrer specifies. It is true that Russell believed that, for instance, propositions about material objects could always be doubted; he also believed that even though propositions about one's beliefs about one's own percepts were "self-evident," they were nevertheless not "certain" (only less doubtful than any others and the foundation on which other inferentially related propositions depended).⁸ In short, Russell only professed a "methodological doubt" in order, precisely, to lead on to some form of foundationalism, somewhat more tentative, it must be said, than Descartes' original view or that of other more recent advocates (for instance, Chisholm).

Lehrer's claim is an interesting one, which, though it raises difficulties of its own, serves rather nicely to focus the dialectical importance of skepticism and the options for resisting skepticism. As Lehrer puts it:

The chance that we are wrong about the truth value of any contingent statement is the probability that we are wrong. The probability that we are wrong is a property, or feature, or propensity of the world that manifests itself in error. This propensity is explained in terms of human conception which continually shifts as a result of our attempt to apply concepts in order to facilitate our goals and objectives.⁹

⁴ Moore gives the relevant references.

⁵ Keith Lehrer, "Skepticism and Conceptual Change" in Roderick M. Chisholm and Robert J. Swartz (eds.), *Empirical Knowledge* (Englewood Cliffs, 1973), p. 47. This seems to be the first appearance of Lehrer's paper. The thrust of the paper conforms quite closely, however, with Lehrer's *Knowledge* (Oxford, 1974).

⁶ *Knowledge*, *op. cit.*, p. 238.

⁷ *Ibid.*, p. 238.

⁸ For instance, in "Philosophic Doubts," *Philosophy* (New York, 1927). This also appeared under the title, *An Outline of Philosophy* (London, 1927). For a full discussion, cf. Roderick M. Chisholm, "Russell on the Foundations of Empirical Knowledge" in Paul Arthur Schilpp (ed.), *The Philosophy of Bertrand Russell* (New York, 1951). It is not unimportant to notice Chisholm's connection between the two issues (skepticism and foundationalism) since this is essentially the pivot of his own *Theory of Knowledge* (Englewood Cliffs, 1966) as well as the motivation for the structure of the reader, *Empirical Knowledge*, *op. cit.*

⁹ "Skepticism and Conceptual Change," *op. cit.*, p. 50.

To locate the strategic point of Lehrer's claim—possibly in a way that we would regard as either too narrow or even not rightly in accord with his intention—one may say that what Lehrer has done is to interpret the skeptical thesis in accord with such views about diachronic changes in our conceptual scheme as have been pressed by Wilfrid Sellars, Thomas Kuhn, and W. V. Quine¹⁰ against the implicit resistance to the force of *this* sort of skepticism entailed in so-called foundationalism (what Lehrer calls the Foundation Theory), most compellingly championed in our own time by Roderick Chisholm.¹¹ One difficulty with Lehrer's view is that he wishes to restrict the thesis to contingent statements, to disallow its being applied to logical truths: the trouble is that if, as he claims, "any concept, even mental ones, or the concept of existence, may be retired from conceptual service and be replaced by other concepts better suited to the job [intended],"¹² then, either the non-fixity of the analytic/synthetic distinction would undermine the restriction or else Lehrer would not be able to base his thesis on "standard conceptions of probability" (which would assign logical truths a probability of one) or else a lacuna in the argument remains still to be filled.¹³

The most important feature of the thesis, however, lies in its focusing a confrontation between skepticism and those doctrines that would avoid skepticism—and, because of that, its focusing an even more important confrontation between the varieties of resisting doctrines, in particular, foundationalism and non-foundationalist solutions. Lehrer, of course, rejects both skepticism and foundationalism and advocates instead what he regards as a coherence theory. In short, he advocates "a theory of knowledge without certainty": hence, he embraces the Moorean form of skepticism (pressed in terms of the implications of conceptual change), regards it as benign enough, and concludes

that we need not accept "the deep skeptical conclusion of universal ignorance."¹⁴

The dialectical possibilities are now clearer. The traditional form of skepticism—universal ignorance (which Russell does not advocate and which Moore does not even formulate as a serious contender)—has been, equally traditionally, offset by some form of foundationalism. Lehrer's characterization is convenient. Foundation theories, usually versions of empiricism, may be said to meet the following conditions: "The first is that a basic statement must be self-justified and must not be justified by any non-basic belief. Second, a basic belief must either be irrefutable, or, if refutable at all, it must only be refutable by other basic beliefs. Third, beliefs must be such that all other beliefs that are justified or refuted are justified or refuted by basic beliefs."¹⁵ Hence, foundationalism requires a demarcation between basic and non-basic beliefs (or basic and non-basic knowledge), where basic beliefs are characterized by indubitability or infallibility or the like or where the possibility of error is radically circumscribed in contrast to the fallibility of non-basic beliefs. The details are not important at this point, though it is helpful to notice that Moore's position is somewhat anomalous, since Moore maintains that we may be certain regarding commonsense beliefs but never explains the full relationship between such beliefs and sense data—with respect to which he seems to hold a foundationalist view.¹⁶

What is important is this. Foundationalism has appeared as a promising theory, in spite of difficulties about self-evident or "self-presenting" contingent truths, because it was felt that the thesis of universal ignorance would otherwise be irresistible. It is precisely the distinction of Lehrer's account, apart from questions one may wish to raise about its detailed claims, that it attempts to provide what Lehrer terms a coherence alternative to founda-

¹⁰ Cf. Wilfrid Sellars, "Philosophy and the Scientific Image of Man," *Science, Perception and Reality* (London, 1963); Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1970, 2nd ed. enlarged); W. V. Quine, *Word and Object* (Cambridge, 1960).

¹¹ Roderick M. Chisholm, *Perceiving: A Philosophical Study* (Ithaca, 1957); and *Theory of Knowledge*, *op. cit.*

¹² "Skepticism and Conceptual Change," *op. cit.*, p. 51. Compare *Knowledge*, *op. cit.*, p. 81, which is noticeably inconclusive.

¹³ Actually, Lehrer seems to have confused the probability of logical truths and their contradictories and the probability of one's knowing a logical truth to be a logical truth; cf. *ibid.*, p. 51. Thus, Lehrer wrongly affirms that "one cannot argue that there is always some probability that one is *wrong* no matter what sort of *statement* one thinks is true without abandoning standard conceptions of probability" (*italics added*).

¹⁴ *Knowledge*, *op. cit.*, p. 239.

¹⁵ *Ibid.*, p. 76.

¹⁶ Norman Malcolm's *Knowledge and Certainty* (Englewood Cliffs, 1963) may be cited also as a Moorean variant of the commonsense view, without foundationalist assumptions. Cf. however Joseph Margolis, "Malcolm on Knowledge," *Logique et Analyse*, vol. 37 (1967), pp. 192–199.

tionalism—at the same time that Lehrer embraces both what Moore takes skepticism to be and rejects traditional skepticism. That is, traditional skeptics have taken the Moorean sort of concession to lead irresistibly to their own position; Lehrer attempts to demonstrate how to avoid this. But Lehrer's own solution is merely one of a battery of possible alternatives to foundationalism, all of which may be somewhat more compendiously characterized as forms of *pragmatism*. This may seem surprising and even misleading. But the sense of pragmatism here intended is such that it could not be captured by any one familiar figure: it is a philosophical orientation sufficiently resilient to accommodate such diverse thinkers as the following: Charles Sanders Peirce, William James, John Dewey, C. I. Lewis, W. V. Quine, Wilfrid Sellars, Thomas Kuhn, Paul Feyerabend—even the Wittgenstein of the *Investigations*, also Gilbert Ryle and J. L. Austin—and, of course, such more recent authors as Lehrer and Gilbert Harman.¹⁷ It would exclude such thinkers as Moore, Russell, Rudolf Carnap, A. J. Ayer through a number of phases, and (most relevantly in our own time) Chisholm, since all of these thinkers have in one way or another insisted on some form of foundationalism; it would also exclude advocates of the new rationalism, Noam Chomsky for instance,¹⁸ though the standard context for disputing the skeptical thesis is empiricism.

There are a number of themes that may be termed pragmatic or pragmatist that are not centrally concerned with our present issue. Nevertheless, it is not unhelpful to specify the characteristic pragmatic doctrines, so that one may see how a variety of philosophical inquiries dovetail with our present issue. These appear to include the following: first of all, the denial of any foundation theory of knowledge, either in terms of self-evident percepts or *sensa* or the like or in terms of self-evident truths on which the body of science depends and must depend; secondly, the thesis that human inquiry is continuous with, and develops out of, the biological and pre-cognitive interaction between organism and environment and that a theory of science must account for that continuity; thirdly, the affirmation of some form of empirical realism, to the extent at

least that human organisms are perceptually in contact with the external world, whatever the internal conditions on which their sentient capacities depend—though of such a sort that contact is inextricably qualified by their conceptual schemes and *vice versa*. With these distinctions in mind, it is obvious that Lehrer's account is intended to be pragmatist in nature. It may be thought that Lehrer's theory courts incoherence insofar as he adopts Moorean skepticism on the basis of diachronic conceptual change. But it must be remembered that Lehrer uses the thesis only to insure the impossibility of certainty regarding knowledge, not in a way that threatens the truth and falsity of anything believed to be true. Coherence rightly concerns the latter issue alone.

There are, however, certain interesting limitations in Lehrer's proposal, that illuminate the strategies available to non-foundationalist opponents of skepticism. Lehrer, as was said, adheres to a coherence theory. The label is easily misunderstood since we are accustomed to speak of coherence and correspondence theories of *truth*. Lehrer in fact attempts to show that so-called correspondence and coherence theories of truth "must either be rejected as inadequate or else . . . reduce to the elimination theory [of truth]."¹⁹ By the elimination theory, Lehrer means that the conditions of truth may be specified jointly by the following schemata:

(AT) It is true that *p* if and only if *p*,

which involves "the absolute conception of truth," in that truth is not relativized to any language; and

(CAm) *S* is true in *L* and *S* means in *L* that *p* if and only if *S* is a sentence in *L* meaning that *p* and *p*,

which involves "the semantic conception of truth," since reference to a language *L* obtains.²⁰ The theory is an elimination theory because, on the argument, reference to truth may be eliminated by the joint use of the right-hand equivalents. A standard objection to Lehrer's view, apparently, is that "a satisfactory theory of truth is one that tells us *why* a true sentence is true [and] the elimination theory merely tells us when a sentence *S* is true—crudely, *S* is true if and only if *p*—but it does not

¹⁷ Cf. Gilbert Harman, *Thought* (Princeton, 1973), whose solution to the problem of knowledge Lehrer opposes, but within the framework that I am calling pragmatic. Cf. for instance Joseph Margolis, "The Relevance of Dewey's Epistemology" in Stephen Cahn (ed.), *New Studies in the Philosophy of John Dewey* (University Press of New England, forthcoming).

¹⁸ Cf. Joseph Margolis, "Mastering a Natural Language: Rationalists vs. Empiricists," *Diogenes*, vol. 84 (1973), pp. 41–57.

¹⁹ *Knowledge*, op. cit., p. 47.

²⁰ *Ibid.*, Ch. 2.

tell why sentence *S* is true.”²¹ But Lehrer demonstrates, on his own view, that “the analysis of knowledge is one thing, and the rules for deciding whether one knows are quite another. The latter are conditions of justification for knowledge claims.”²²

The trouble is that it is not so easy to gain agreement that we can analyze the *nature* of knowledge without, at the same time, providing an account of the *criteria* (on Lehrer’s view, “the rules for deciding whether one knows”) for proper ascriptions of knowledge or *vice versa*. The reason is instructive. The notions of truth and knowledge are quite different. It is possible in principle to theorize about *what it means* to say that a statement or proposition is true, without supposing that we are in a position to provide the adequate criteria of truth. Presumably, this is precisely what the classical correspondence and coherence theories of truth had attempted to provide (with what success, may be questioned), because standard defenses of such theories (even on Lehrer’s view)

never attempt to derive criteria of truth from their analyses. In fact, this is normally though not necessarily convincingly, taken to be a sign of their eliminability or pointlessness. Now, in the more-or-less standard *analysis* of knowledge, admitting the version provisionally advanced by Lehrer:

- (i) If *S* knows that *p*, then it is true that *p*;
- (ii) If *S* knows that *p*, then *S* believes that *p*;
- (iii) If *S* knows that *p*, then *S* is completely justified in believing that *p*.²³

it is clear that (i) alone makes no reference, in the apodosis, to attributes of a cognitive agent. Hence, even if (i) were allowed in the analysis of knowledge—which the elimination theory is alleged to make otiose—it could never serve to yield *criteria of knowledge* (as the others presumably could) unless we possessed (apart from a theory of the nature of truth) a defensible account of the *criteria of, or the conditions for rightly ascribing, or for testing for the presence of, truth*. Lehrer rightly remarks that a theory which tells us “how we *determine* that a

²¹ *Ibid.*, p. 42. The most sustained account of the coherence theory of truth is provided by Nicholas Rescher, *The Coherence Theory of Truth* (Oxford, 1973). But Rescher’s theory cannot be regarded as an orthodox coherence account (p. 262) and it explicitly insists on the non-eliminability of truth. For one thing, pursuing the same issue that Lehrer raises, Rescher emphasizes (correctly) that Tarski’s theory provides neither an explicit definition of truth nor a criterion of truth; he stresses that these may reasonably be required and that “any theory of truth—correspondentist or otherwise—can accept the Tarski condition” (p. 19, note 29). Secondly, regarding eliminability, particularly with respect to the views of F. B. Ramsey and A. J. Ayer, Rescher observes: “the thesis that ‘is true’ is *assertively redundant* because ‘*P* is true’ provides the same information as the mere assertion of ‘*P*’ itself cuts both ways. It argues also that declarations are *assertively expansive* with respect to ‘is true,’ because the mere declaration of ‘*P*’ amounts to the declaration of ‘*P* is true.’ Thus assertive redundancy, rather than showing ‘is true’ to be inane and dispensable, can be construed to show that it is important because ubiquitous, and represents an omnipresent—if tacit—feature of assertions in general” (p. 7, note 14). Thirdly, Rescher openly concedes that truth cannot be *defined* in terms of coherence; he appears to favor a correspondence theory of truth, which certainly entails the ineliminability of truth (pp. 9–10, 12, 16, 23, 27–28, 262). He says explicitly that “the position we shall defend supposes that coherence is not the *meaning* of truth in the context of factual claims, but its *arbiter* (to use F. H. Bradley’s well-chosen word)” (p. 12). But finally, not only does he concede the importance of correspondence elements in an account of truth, he also insists that “the justification of the coherence methodology itself can be neither deductive nor inductive in character . . . [but] pragmatic” (p. 232). Here, he rejects classical conceptions of coherence (Ch. 2) and elaborates the required form of a pragmatic justification of the coherence criterion in a way that makes it extremely difficult to draw a sharp distinction between coherentist and non-coherentist accounts (Ch. 10). He characterizes his own thesis as “criterion pragmatism,” not “proposition-pragmatism” (p. 237)! It is not part of our present concern to examine the force of Rescher’s thesis. But it may be noted that his purported justification of coherence depends (a) on the tenability and relevance of distinguishing between cognitive or theoretical concerns (truths and facts) and practical or affective concerns (satisfactory guides to action) (pp. 239–242); and (b) on the tenability and relevance of distinguishing between the practical validation of particular acts and of general rules to which particular acts conform (paralleling the distinction between act- and rule-utilitarianism) (pp. 242–244). But, on the face of it, to confirm that actions are in some sense satisfactory appears to entail that facts regarding their purported satisfactoriness must be independently established, which threatens the proposal with circularity: the theoretical and the practical appear unable to be relevantly distinguished. Secondly, the prospect that the validation of rules guiding action and the validation of particular actions themselves may be collapsed into one another constitute another genuine threat to the proposal; the parallel issue regarding utilitarianism (Rescher invokes the comparison) has been effectively explored by David Lyons, *Forms and Limits of Utilitarianism* (Oxford, 1964). Rescher himself poses the most difficult objections and clearly regards his defense as the sketch of a promising strategy (pp. 244–262). It depends essentially on the tenability of distinguishing questions of truth and presumptiveness or plausibility in a sufficiently strong way (pp. 249–251, 255–256); but questions of plausibility may, after all, be parasitic on questions of truth. The admission of a pragmatic justification, therefore, shows at the very least that we must understand coherence in a novel way or the sufficiency of the coherentist account must be conceded to be as yet uncertain. Lehrer’s argument fails to come to grips with these difficulties.

²² *Ibid.*, p. 48.

²³ *Ibid.*, Ch. 1.

sentence is true or *justify* the claim to know that it is true is not a theory of truth at all but a theory of justification."²⁴

Well and good. But then, the elimination theory of truth not only does not (bother to) explain the nature of truth but does not (bother to) provide us with criteria or rules for the ascription of truth; and the question remains whether, in spite of the eliminative maneuver, either reference to truth or reference to whatever it is that establishes that *p*, can fail to be provided in an account of the *justification of ascriptions of knowledge*. From this point of view, Lehrer may be said to offer a coherence theory of belief or, more perspicuously, a coherence theory of the justification of belief: here, he is concerned with the conditions or criteria on which belief is justified—which is important. As he himself says, "As a coherence theory, our theory affirms that a man is completely justified in believing that *p* if and only if the statement that *p* coheres with a system of kind *k*. The system of kind *k* is a system of beliefs. It contains statements describing the beliefs of *S* which we call a doxastic system of *S* [that is, the set of statements that describe what *S* believes, as '*S* believes that *p*,' '*S* believes that *q*'] . . . [corrected]. A veracious man seeks to believe only what is true and all that is true. A corrected doxastic system of a veracious man must describe a consistent set of beliefs to make it logically possible for him to avoid erroneous belief."²⁵ The trouble is that *this* characterization contains an explicit reference to truth, albeit attenuated: the coherence theory of belief applies only to corrected doxastic systems, but such systems are doxastic systems corrected to conform with the intentions of a veracious man; and if we lack criteria or rules for *determining* truth or what it is to believe only what is true, there would appear to be no operative basis for distinguishing the veracious man from the ordinary man, except in terms of his mere intentions or say-so—which can hardly be supposed to bear decisively on ascriptions of knowledge.

Lehrer actually enlarges the telltale qualification noted here. Corrected doxastic systems simply involve the intentions of a veracious man and provide for the logical possibility of avoiding erroneous belief—that is, believing what is false. But Lehrer himself admits that "it remains a

logical possibility that the beliefs of a man about the world described within his corrected doxastic system could be entirely erroneous;"²⁶ so that "completely justified true belief" that rests on a corrected doxastic system is not knowledge.²⁷ Actually, Lehrer's final resolution comes to this:

a man knows that *p* only if he is completely justified in believing that *p* in the verific alternative to the corrected doxastic system. The verific alternative to a corrected doxastic system is one in which all statements describing false beliefs in the corrected doxastic alternative are replaced with statements describing true beliefs. If the statement that *S* believes that it is *p* is in the corrected doxastic system of *S*, and it is false that *p*, then in the verific alternative the statement is replaced with the statement that *S* believes that it is not the case that *p*. The verific alternative is, therefore, a description of what *S* would believe were he to replace each erroneous belief with the contradictory of what he actually believes. If a man is completely justified in believing that *p* in his corrected doxastic system and he is also completely justified in believing that *p* in the verific alternative of that system, then his being completely justified does not depend on, nor is it defeated by, any false statement or belief.²⁸

But then, clearly, both the corrected doxastic system and the verific alternative can be specified only on the condition that true and false beliefs can be independently specified; in fact, although the verific alternative is formulated in terms of a corrected doxastic system, a corrected doxastic system (if it is to have more than the barest intentional force) can be specified only if the verific alternative can be independently specified.

In short, Lehrer recognizes the inherent limitation of a coherence theory of justified belief; and he himself introduces as a necessary condition for justified ascriptions of knowledge a constraint that cannot be relevantly captured by coherence itself. The elimination theory of truth, therefore, is either an illusion or else has no operative force as far as the problem of the conditions that justify construing belief as knowledge is concerned. The upshot is that, although Lehrer's general strategy against skepticism is entirely reasonable, his particular argument (the coherence theory, a version of what has here been termed pragmatism) utterly fails to unseat the skeptic. As already noted, Lehrer correctly observes that the skeptic holds that if there

²⁴ *Ibid.*, p. 47.

²⁵ *Ibid.*, p. 212.

²⁶ *Ibid.*, p. 213.

²⁷ *Ibid.*, p. 214.

²⁸ *Ibid.*, pp. 234-235.

is "some chance" that *S* is wrong in his belief that *p*, then *S* does not know that *p*. But Lehrer claims that, on the analysis provided (sketched just above), "the premiss is unavailable." "It does not follow," Lehrer says, "from the premiss that there is some chance that *S* is incorrect in his belief that *p*, that *p* is not true, or that *S* does not believe that *p*, or that *S* is not completely justified in believing that *p*, or that *S* is not completely justified in believing that *p* in his verific alternative."²⁹ True enough. None of these conclusions follow. But in order to *refute* skepticism (the relevant section is actually titled, "A Refutation of Skepticism"), what we need is not only a stalemate regarding the required entailments but a demonstration that the skeptic is unable to interpose any further reasonable or valid premisses from which the skeptical conclusion follows. Lehrer wrongly believes he has succeeded because, apparently, he fails to see that the coherence theory of justification simply does not provide the sufficient conditions of knowledge, that its very plausibility rests on reintroducing the distinction between true and false beliefs in a way that, in effect, ignores the skeptic's challenge. For the admission of the corrected doxastic and verific systems presupposes the very knowledge (of how to determine what is true and false) which the skeptic challenges. "We avoid skepticism," Lehrer claims fairly enough, "by construing a theory of justification without a guarantee of truth."³⁰ But this only shows that knowledge does not entail certainty—once again, against Moore's view (and foundationalism in general). But nothing that Lehrer has said bears on constructing a theory of justification *without a criterion of or rule for determining truth*: on the contrary, his own theory requires it—as well as the genuine refutation of skepticism. It may therefore be that a viable theory of knowledge avoids foundationalism by giving up certainty *and* avoids but does not refute skepticism by providing an operative criterion of truth. Lehrer, in effect, hobbles his own theory by eliminating a condition he himself, in effect, recognizes to be necessary to any viable theory of justification.

In effect, this is the point of insisting that an analysis of knowledge is impossible without the provision of adequate criteria of knowledge, and

vice versa. Chisholm had (we now see, rightly) insisted that there is no way of answering either question ("What do we know?," "How are we to decide, in any particular case, *whether* we know?") without the other.³¹ The reason is plain: knowledge is a privileged condition of *ourselves* (whereas truth is not). How could we possibly understand what it is to know unless we were in a position to specify the conditions on which knowledge could rightly be ascribed to ourselves? Hence, we see why it is that foundationalists both insist on the connection and avoid skepticism by introducing some form of basic or self-evident knowledge.

It is extremely suggestive that the first entry of Chisholm and Swartz's new anthology³² is Leonard Nelson's "The Impossibility of the 'Theory of Knowledge'." Because the paper introduces in effect the rationale for Chisholm's *Theory of Knowledge*. This is not to say that Chisholm subscribes to Nelson's neo-Kantian arguments regarding so-called "nonintuitive immediate cognition" (or synthetic *a priori* metaphysical judgments) or even the details of Nelson's version of foundationalism. But Nelson raises the spectre of skepticism—that is, the threat of ignorance—precisely in the same way that Chisholm does, by noticing that the questions of the nature and criteria of knowledge cannot be pursued in any way that assigns logical priority to the one or the other. This is the point of Nelson's title. But, of course, we are saved from skepticism, introduced by way of the "theory of knowledge" itself, by acknowledging what Nelson regards as the "psychological facts."³³ Here, there is an extraordinary (and, one may almost say, outrageous) similarity to be drawn to Moore's way of proceeding. Moore confutes Russell's skeptical challenge by *announcing* (by doing no more than announcing) that he has certain knowledge of certain common-sense matters; and Nelson baldly maintains that the "theory of knowledge" simply "contradicts the facts of inner experience," that we just *have* certain cognitions regarding which there is no "possibility of error."³⁴

But that is perhaps a quibble. The more important feature of Nelson's argument lies in the clue it inadvertently provides about a distinctive maneuver among foundationalists—which exposes

²⁹ *Ibid.*, p. 238f.

³⁰ *Ibid.*, p. 241.

³¹ *Theory of Knowledge*, *op. cit.*, p. 56.

³² *Empirical Knowledge*, *op. cit.*

³³ "The Impossibility of the 'Theory of Knowledge'," *op. cit.*, p. 8.

³⁴ *Ibid.*, pp. 8, 9.

the vulnerability of their essential claim. The following passage repays close attention:

To convince ourselves of the existence of cognitions that are not judgments, we need only consider any intuition at all, such as an ordinary sensory perception. For example, I have a sensory perception of the sheet of paper that lies here on the table before me. This perception is, first of all, a cognition, not merely a problematic notion. The existential assertion that is an element of this cognition is, however, not a judgment. To be sure, I can also render in a judgment the same circumstances that I here cognize through the perception; but when I judge that a piece of paper is lying before me on the table, that is an altogether different sort of cognition from the perception of this situation. I need concepts for the judgment, e.g., the concept "table," the concept "paper," etc. I connect these concepts in a certain manner and assert that objective reality pertains to this combination of concepts. Perception, on the other hand, has no need of any concepts nor of any problematic notion of its objects whatsoever; rather, it is itself an originally assertoric notion—is, in other words, an *immediate* cognition.³⁵

This is an amazing claim. For one thing, Nelson obviously holds that these "cognitions" have a certain assertoric force (in fact, have the force of a truth claim without making any claim), are not judgments though they have such force, and are valid (as providing knowledge) without the mediation of concepts! For another, such "cognitions" are inner mental states of some sort. Now, it can be fairly argued that no cognition and no mental state that, on a theory, qualifies or counts as knowledge can possibly be relevantly specified except by articulating its content propositionally. Thus, *S* fears that it will rain; *S* thinks that today is Tuesday; *S* sees that there is a piece of paper lying before him. But it is impossible to concede that cognitions are propositionally detailed without also conceding that they count as *judgments* (or involve judgments, if we think of knowledge as a capacity)—acknowledging some "assertoric" force that is the analogue of the illocutionary force of actual speech acts.³⁶ And it is impossible to concede this without also conceding that cognitions (or judgments that are cognitions) involve the use of *concepts*, that is, the relevant capacity to understand (in some sense, not necessarily linguistically, as with the higher

animals) that what may be affirmed by the use of propositions is the case.

The reason this connection is important is simply that if Lehrer's claim about the "chance" of error be admitted, then the certitude of such basic "cognitions" is effectively undermined. Only if immediate cognitions were not (or did not involve the use of) judgments, did not involve the use of concepts, could the thesis of self-evidence, indubitability, complete certitude, and the rest have any inning at all. Nelson's thesis gives the appearance of a viable foundationalist claim precisely because Nelson himself insists that immediate cognitions are not judgments. They are, as he says, psychological or inner mental states. *But* if they are mere states lacking "assertoric" properties, then there is no prospect that they can be the appropriate analogues of judgments. It is idle, therefore, for Nelson to raise as a puzzle "the possibility of error"; the simple reason Nelson's would-be "cognitions" cannot be in error is that Nelson insists, contradictorily, that they both have "assertoric" force and yet are not (or do not involve) judgments at all. The "chance" of error—in Lehrer's sense—immediately obtains, once the correction is made. So the surface plausibility of Nelson's maneuver rests squarely on his having conflated cognition as (or as entailing) a judgment that succeeds in a certain appropriate way (explicit or implicit, linguistically formulated or not so formulated) and "cognition" as some sort of psychological state that is not (or does not involve) a judgment. Knowledge, in fact, cannot be straightforwardly regarded as a psychological state, in the manner of belief, and must be analyzed in a way that is utterly different from belief. To put the matter squarely: beliefs are psychological states but knowledge is a certain *status* assigned to beliefs or beliefs and certain capacities and skills.³⁷

The foregoing, then, illuminates in an extremely perspicuous way the fundamental difficulty of Chisholm's foundationalism—which, without a doubt, is the most sophisticated version of the doctrine that is currently available. For in speaking about the "directly evident"—which, on Chisholm's analogue of Nelson's worry about an infinite and vicious regress, provides us (foundationally) with "a proper stopping place" against the Socratic

³⁵ *Ibid.*, pp. 8–9.

³⁶ Cf. Zeno Vendler, *Res Cogitans* (Ithaca, 1972), Ch. 3. I have grave reservations about Vendler's general thesis, but not at this point. These are formulated in an as yet unpublished paper, "Knowledge and Belief; Facts and Propositions."

³⁷ Cf. *Knowledge and Existence*, *op. cit.*, Ch. 1.

questions of justification—Chisholm offers the following formula:

What justifies me in counting it as evident that *a* is *F* is simply the fact that *a* is *F*.³⁸

Chisholm's clarification is instructive. He borrows a term from Meinong and says that "what is directly evident to a man is always some state of affairs that 'presents itself to him.' Thus, my believing that Socrates is mortal is a state of affairs that is 'self-presenting' to me. If I do believe that Socrates is mortal, then, *ipso facto*, it is evident to me that I believe that Socrates is mortal; the state of affairs is 'apprehended through itself'."³⁹ Chisholm *does* question Nelson's formula. For he takes it that Nelson has not quite sorted out the distinction between immediate experience and the justification for counting immediate experience as evident.⁴⁰ Here, he finds Leibniz's formulation more to his liking ("primary truths of fact"). But his point is that, rather in the way Nelson views the matter, the psychological experience—or, more accurately, the fact of having a certain psychological experience—serves to justify my counting it as evident that *a* is *F* because "the state of affair" is "self-presenting" or "directly evident" or, Leibniz's phrase, such that "there is no mediation between the understanding and its objects."⁴¹ Nevertheless, apart from plausible but piecemeal challenges to Chisholm's particular instances⁴² or a general attack on self-evidence and indubitability,⁴³ there can be no question that Chisholm's formulation is open to challenge on the basis of Lehrer's conception of the "chance" of being mistaken, that is, on the basis of the significance of diachronic conceptual changes. For, either Chisholm adheres to Nelson's view of "cognitions" (which seems not to be the case) or else Chisholm construes the apparent cognition as involving a judgment ("What justifies me in counting . . ."); but if he favors the latter alternative, then Lehrer's claim has force. Hence,

foundationalism (that is, Chisholm's variety, though not necessarily Russell's) appears to be subject to a fatal weakness. Needless to say, any theory of knowledge that subscribes to a coherence conception of truth is, to that extent, incompatible with (any form of) foundationalism.⁴⁴

What we see, then, is that skepticism may not be demonstrably refuted—neither by the foundationalist countermove nor by that sort of pragmatic maneuver that Lehrer terms a coherence theory of justification. If we may concede Moorean skepticism without affirming total ignorance, then it may well be that we can construct a reasonable theory of knowledge that (i) does not require foundationalism; and (ii) *does not entail the refutation of skepticism*. We may, therefore, avail ourselves of the point Lehrer correctly presses (without drawing Lehrer's further conclusion about the refutation of skepticism), viz., that the skeptic himself has yet to prove that "if there is some chance that *S* is incorrect in his belief that *p*, then *S* does not know that *p*." We do not have a stalemate, here, because the skeptical consequence does not follow from Moorean skepticism; and we do not need a refutation, since the plausibility of ascribing some knowledge is patently linked with every distinctly human effort and strengthened by the very attempt to explore the tenability of the skeptical claim itself. We do not need to *demonstrate* that we have knowledge: that would only lead to Nelson's predicament again; we merely need to *theorize* about the nature and criteria of what we are prepared to count as knowledge—not being obliged to reject it entirely.⁴⁵ Consequently, what we see is that a characteristically powerful theme in contemporary philosophy—that our conceptual scheme changes diachronically—is directly opposed to foundationalism and directly inclined to favor what has here been termed pragmatism. Theories of such a sort, therefore, are opposed to skepticism but need not be designed to refute it.

Temple University

Received February 24, 1976

³⁸ *Theory of Knowledge*, *op. cit.*, p. 28.

³⁹ *Ibid.*, p. 28. Here, he cites in a footnote, apparently approvingly, Meinong, Brentano, Russell, Ledger Wood, and C. J. Ducasse.

⁴⁰ *Ibid.*, pp. 27–28.

⁴¹ *New Essays Concerning Human Understanding*, Bk. IV, Ch. 9; cited by Chisholm.

⁴² Cf. Joseph Margolis, "Directly Evident," *Theoria*, vol. 34 (1968), pp. 102–116.

⁴³ Cf. Joseph Margolis, "Indubitability, Self-intimating States, and Logically Privileged Access," *The Journal of Philosophy* vol. 67 (1970), pp. 918–931.

⁴⁴ Cf. Rescher, *op. cit.*, Ch. 13.

⁴⁵ Rescher comes to a similar conclusion, *ibid.*, pp. 330–331.

V. THREE SCEPTICAL THESES IN ETHICS

FRANK SNARE

THERE are three meta-ethical theses commonly held by sceptics which are thought to be closely related. Rarely, however, do sceptics—or their critics—bother to explain what the relationships are. In fact it is not unusual for both the sceptic and his critic to shift from one to the other in discussion without explanation or apology. Since such shifts are made easier by the fact that these theses are not always stated with any great precision, we do well to begin by trying to formulate each of the three claims. The first thesis, non-cognitivism, is fairly straightforward. This is the claim that ethical judgments are not the sort of thing to be true or false. They are not cognitive. In contrast there are a couple catchphrases of scepticism in ethics which need more explication: the slogan “*Ought* cannot be derived from *is*” and the reproach “Naturalistic fallacy!”. It is worthwhile getting clear just what claims might be behind these two catchphrases and how they relate to the non-cognitivist thesis. The place to begin is Frankena’s article “The Naturalistic Fallacy” which remains one of the best discussions of G. E. Moore’s use of this notion in *Principia Ethica*.¹ Frankena distinguishes the following claims:

- a. Ethical propositions are not deducible from non-ethical ones.
- b. Ethical characteristics are not definable in terms of non-ethical ones.
- c. Ethical characteristics are different in kind from non-ethical ones.

These formulations are quite appropriate if we wish to discuss Moore who, although he was definitely not a value sceptic, did want to force a gap of sorts between *is* and *ought* and soundly condemned a “naturalistic fallacy.” However, the formulations are in a certain respect inappropriate for the sceptic about ethics. For example, if one thinks of a proposition as that sort of thing which alone is capable of being true or false, the sceptic should deny that there are any ethical propositions. Likewise, if a characteristic is something like an ordinary property, the sceptic perhaps does best to

deny there are ethical characteristics. Now this might seem a very good reason for rejecting all three of Frankena’s formulations in favour of the simple non-cognitivist thesis that no ethical judgment is a cognitive judgment. No one disputes that moral judgments are made as they dispute whether ethical propositions exist. Nevertheless it would be rash at this point to reject (ε), (b), or (c). There are several ways out of the present difficulty. One way out would be to drop the term “proposition” and speak instead of judgments, sentences, speech acts, claims, or whatever. Again, the sceptic can use “term” or “predicate” instead of “characteristic.” A second way out would be to adopt a suitably broad sense of “proposition” and “characteristic.” For example, we might think of a proposition as whatever it is that is referred to when we fill in the blank in “*X* believes that _____.” Thus we may speak of the proposition that Jones has an obligation to return the book to Smith since we can say that *X* believes that Jones has an obligation to return the book to Smith. There is no presupposition that such a “proposition” has to be true or false. Propositions are simply whatever may be the objects of propositional attitudes. Now I think it is important to keep open both avenues of escape for the sceptic. I shall keep “proposition” and “characteristic” in Frankena’s formulations with the understanding that they are to be interpreted in an extended sense. However, those who wish may make an appropriate substitution in terms of judgments, sentences, or whatever.

It is the denial of (c) which, in an overly strict sense, is the much denounced naturalistic fallacy. Nevertheless I propose to drop (c) from consideration. It has special relevance in the discussion of Moore but not in the discussion of scepticism in ethics. Although (c) implies (b), it turns out to be more difficult to show (c) than to show (b). This is not surprising since (c) asserts something in particular beyond what (b) asserts. Ethical characteristics are said to belong to an entirely different logical, or metaphysical, order from non-ethical ones and this is the reason the former cannot be

¹ W. K. Frankena, “The Naturalistic Fallacy,” *Mind*, vol. 48 (1939), pp. 464–477.

defined in terms of the latter; thus (b) is the more general sceptical thesis. Furthermore, if Moore's famous open question test can be used at all, it can be used to show (b) is true. So (c) is unimportant for our purposes.

There is another alteration we need to make in Frankena's formulations. It should be noted that the contrast between ethical and non-ethical is appropriate in a discussion of Moore. But in a discussion of ethical scepticism the contrast must be between ethical propositions and cognitive (factual, descriptive) ones. (Likewise the contrast must be between ethical characteristics and descriptive ones.) Moore, of course, only wished to drive a gap between two kinds of facts (ethical and non-ethical) or two kinds of properties (ethical and non-ethical). However, the sceptic would like to drive a gap between ethical judgments and cognitive judgments, i.e. not between two kinds of facts but between facts and a kind of non-fact. Of course this does introduce notions such as "cognitivity" and "descriptivity" which Frankena's formulations wisely avoided, but I do not see how the sceptic can settle for anything less. Thus, while the sceptic may employ some of Moore's arguments and some of his terminology, he should rewrite claims (a) and (b) as:

1. Ethical "propositions" are not deducible from cognitive ones.
2. Ethical "characteristics" are not definable in terms of descriptive ones.

Claim (1) is the *no-ought-from-is* claim. The naturalistic, or defnlist, fallacy is thought to involve one in the denial of (2). Both assert a kind of gap between facts and values. However, a gap is also asserted in the basic claim of non-cognitivism:

(N) No "proposition" is both ethical and cognitive.

It is the task of this paper to investigate the relationships between these three theses and perhaps to determine which thesis is the broadest. We will begin by comparing (N), non-cognitivism, with (1), the *no-ought-from-is* thesis. There is some argument for supposing that (N) and (1) entail each other.

1. WHETHER (N) ENTAILS (1)

It can be argued that (N) entails (1) provided it is allowed that any proposition which follows

from cognitive premisses is also cognitive. For if (1) were false, i.e. if some ethical proposition did follow from cognitive propositions, then it would follow that the ethical conclusion was also cognitive and thus (N) would be false. That is, $\sim(1)$ entails $\sim(N)$, or (N) entails (1). Now the required assumption that cognitive premisses entail cognitive conclusions is plausible under some conceptions of entailment. Entailment as strict implication would be one. Under this view, the premisses entail the conclusion if and only if it is a necessary truth that if the premisses are true, the conclusion is true. Now of course such a relationship can hold only between premisses and conclusions capable of being true or false. Thus there is no question of cognitive premisses entailing a non-cognitive conclusion and our required assumption is saved. Nor does the so-called paradox of strict implication provide us with any further difficulties. For while a contradiction strictly implies every cognitive proposition, it will strictly imply no ethical proposition which is not also cognitive. So it is clear that if a sceptic holds (N) and if he understands "deduce" in (1) in terms of strict implication, he is committed to (1) as well.²

But there is a respect in which the nature of entailment seems to be unduly restricted here. Certainly it is plausible to suppose that there are entailment relationships between commands (which are presumably non-cognitive). Furthermore, I think even a sceptic should allow that there is such a thing as deontic logic or, at any rate, that there are logical relationships between various moral judgments. In fact it might be noted that most sceptics do allow that ethical (and thus non-cognitive) propositions can be deduced from premisses which include cognitive propositions as long as there is at least one ethical (and thus non-cognitive) proposition among them. Sceptics typically allow that we can deduce:

- (i) *a* is pleasant
- (ii) all pleasant things are good
- (iii) *a* is good.

But since such a sceptic insists that (ii) and (iii) are non-cognitive, he is thereby allowing that an entailment relation may hold between propositions some of which are not cognitive. But this example in fact does even more. If the sceptic admits the above entailment, he must surely accept the following entailment as well:

² Cf. Jonathan Bennett, "Entailment," *The Philosophical Review*, vol. 78 (1969), pp. 197-236. A plausible alternative to strict entailment is given in Richard and Valerie Routley, "The Semantics of First-Degree Entailment," *Nous*, vol. 6 (1972), pp. 335-359.

(i) *a* is pleasant

(ii) *a* is good $\vee$ not all pleasant things are good

where the conclusion certainly seems to be ethical, or at least non-cognitive.³ But let me give a second example of an entailment which seems to undermine the assumption that purely cognitive premisses cannot entail anything but cognitive conclusions:

(i) *a* is red

(ii) *a* is red $\vee$ *a* is good.

The conclusion here will probably be considered ethical, or at least non-cognitive by those sceptics who consider any judgment non-cognitive which is in the least way contaminated with ethical notions. But we can say more than this. If a sceptic accepts the disjunctive syllogism, then he will have to say (ii) is non-cognitive.⁴ The argument is this. Since *a is good* is indisputably ethical, the sceptic must say it is non-cognitive. But then he cannot say that (ii) is cognitive since it along with the indisputably cognitive proposition *it is not the case that a is red* entails (via the disjunctive syllogism) the ethical proposition *a is good*. Thus we would have an ethical conclusion from two cognitive premisses. But the alternative is only a little better. For if (ii) is non-cognitive, then our above example shows it to be derived from the cognitive premiss (i).⁵ And this suggests that cognitivity is not inherited via entailment.

In the light of such examples, the sceptic has a choice between two alternatives. On the one hand, he can take the hard-line view that entailment is a relationship which holds only between cognitive premisses (as will be his view if he takes entailment to be strict implication). Under this view cognitivity will be inherited via entailment. Thus, it will be held that (N) entails (1). The other alternative is to take the view that entailment is not restricted to a relationship between cognitive propositions. But in this case it would appear that the two examples we have given will make it difficult for the sceptic to maintain that cognitivity is hereditary over entailment. Thus, he may have to give up the claim that (N) entails (1). Indeed it now seems clear that (1) asserts a bit more than (N) does. (N) allows that, while cognitive premisses might entail (in the

extended sense) ethical conclusions, such conclusions are never cognitive. But (1) disallows the deduction even of a non-cognitive *ought* from an *is*. But perhaps the sceptic will want to hold that what extra (1) asserts is not very important. To show how this might be let us suppose that the above two examples actually show more than has been hitherto claimed. Let us suppose these examples actually show (1) to be false. That is, let it be granted that the two conclusions are not merely non-cognitive but actually ethical. I still do not think a sceptic need find these two entailments very threatening. He can say, first of all, that just because an ethical proposition is entailed, *in the extended sense*, by cognitive propositions does nothing to show it is cognitive. Thus, (N) can still be maintained. But, secondly, I think the sceptic may want to say that the two entailments are not particularly interesting ones in any case. It should be noticed that nothing seems to depend on the actual ethical notions which appear in the conclusions so as to make them ethical. For instance, while *a is red* does entail *a is red $\vee$ a is good*, it also, for the same reasons, entails *a is red $\vee$ a is bad* as well as *a is red or a is indifferent*, not to mention *a is red $\vee$ a is round*. Now of course these various alternative conclusions do not contradict each other, but what is shown is that the ethical notions occurring in the conclusion are not essential to the entailment. And this is peculiar.

Nevertheless, it is too easy for the sceptic simply to say that (1) is not worth maintaining once the inheritance claim has been abandoned. Perhaps the two examples—even assuming the conclusions are ethical—will not be very interesting to the sceptic. But it seems that other conceivable cases should be. For example, if *a is good* can be deduced simply from *a is sought by all men*, then, even if we cannot on this account show *a is good* to be cognitive, it must still be distressing to the sceptic that such a conclusion can be supported in this strong manner by a cognitive premiss which may well be true. The sceptic must not confine himself to too narrow a conception of justification in ethics. Surely it will be a sufficient refutation of scepticism to produce an entailment such as is alleged in the present example. Even if the conclusion cannot be called "true," the truth of the premiss will lend a

³ See David Rynin, "The Autonomy of Morals," *Mind*, vol. 66 (1957), pp. 308-317.

⁴ Some philosophers in order to avoid the "paradoxes" of strict implication have rejected the disjunctive syllogism. See A. R. Anderson and N. D. Belnap, Jr., "Tautological Entailments," *Philosophical Studies*, vol. 13 (1962), pp. 9-24. See also Richard and Valerie Routley, *op. cit.*

⁵ Cf. A. N. Prior, "The Autonomy of Ethics," *Australasian Journal of Philosophy*, vol. 38 (1960), pp. 199-206. Prior is concerned whether certain propositions must be regarded as ethical whereas I am only concerned with whether they must be regarded as non-cognitive by the sceptic.

strong degree of "acceptability" to the conclusion.⁶ It will be inconsistent to assert the truth of the premiss while denying the conclusion (even if such a "denial" does not have a truth value).

So it would seem that, while (N) does not imply (1)—if cognitivity is not inherited—nevertheless it is important for a sceptic to maintain something beyond (N) which is a part, at least, of what (1) asserts. We can show this more clearly by making the following rough distinction. We will call an entailment "interesting" where there is no predicate or proposition in the conclusion such that it can be replaced, in some or all of its occurrences in the conclusion (only) with any other predicate (or proposition) *salva validitate*.⁷ By this criterion our two former entailments will not be "interesting" since whatever proposition we substitute for *a is good* in the one, or whatever predicate we substitute for *good* in the other, will not affect the entailment. In contrast the alleged entailment from *a is sought by all men* to *a is good* would probably be interesting since those who defend such an entailment do not think that they can substitute *bad* for *good* in the conclusion (only) and preserve the entailment. Now the point is that, even if the sceptic is not concerned to deny that there are "uninteresting" entailments, he ought to be concerned if there are "interesting" entailments from cognitive premisses to ethical conclusions even if such conclusions are not, strictly speaking, cognitive. It should distress the sceptic if non-cognitive ethical propositions can be supported *in this way* by cognitive premisses.

Now I think the reply of the sceptic here would be to admit that he would be distressed by any such "interesting" entailment and that it is part of his position to deny that there are any. However, I think it also simplifies matters if the sceptic insists that any proposition entailed by cognitive premisses in the "interesting" way should be properly classified as cognitive. We may suspect that many claims within science and mathematics have no better claim for being called cognitive, e.g. where they can be shown true only by deriving them from true premisses or shown false only by deriving their negation from true premisses. Of course, it is easy to see why saying this is not adequate in the case of "uninteresting" entailments. In such cases some of the terms or propositional components are not

essential to the derivation so that if these terms are themselves non-sensical, non-cognitive, or inadequate in some way, the whole conclusion may turn out to be non-cognitive even though the basis for entailment remains the same. But where every term and component of the conclusion cannot be arbitrarily altered without bringing the entailment into question, this will not be our problem. What this means is that, even in the face of our two former examples, the sceptic may continue to assert (1) and claim that (N) entails (1), provided he understands "deducible" in (1) to mean, not entailed, but something like entailed in the "interesting" manner. (N) will then entail (1), not because cognitivity is preserved in all entailments, but because it does not hurt to say cognitivity is preserved in all "interesting" entailments.

In conclusion, I want to say the sceptic has two plausible choices. First, he can understand "deducible" in (1) to mean entailed where this is taken to be a relation restricted to cognitive propositions. In this event (N) will entail (1) for a rather trivial reason, i.e. the sceptic will deny that *ought* can be "deduced" from *is* because, as he sees it, there can be no real deduction where the alleged conclusion is not cognitive. (N) will become the important thesis and, although (1) will follow, it will seem to have no independent interest in its own right. Secondly, the sceptic can allow that entailment is not restricted to a relation between cognitive propositions but nevertheless restrict "deduction" in (1) to "interesting" entailments. Perhaps the crude criterion I suggested for distinguishing "interesting" from "uninteresting" entailments will be found to be inadequate and will have to be altered in the light of further examples. In any case, such a sceptic will draw the line somewhere arguing, at the same time, that there is no reason to deny that cognitivity is transferred over "interesting" entailments. Thus by this position, as well as by the former, (N) entails (1) and cognitivity is preserved over entailment—or at least over "interesting" entailments.

2. WHETHER (1) ENTAILS (N)

On the other side, the argument that (1) entails (N) is quite conclusive. This is so since $\sim$ (N) entails

⁶ I owe this objection to Stephen Voss.

⁷ See the discussion of "contingent vacuity" in Prior, *op. cit.* and also the criticism by David R. Kurtzman in "'Is,' 'Ought,' and the Autonomy of Ethics," *The Philosophical Review*, vol. 79 (1970), pp. 493–509. I do not want to suggest that in an "uninteresting" entailment the entailment, the conclusion, or any term in the conclusion is "vacuous" in any ordinary sense. Furthermore, my discussion is restricted to the issue of the inheritance of cognitivity and then only under a certain view of entailment.

$\sim(1)$, i.e. if a proposition were both ethical and cognitive then—because a (cognitive) proposition always entails itself—an ethical proposition would entail a cognitive one. Now it is instructive to note the following: (N) follows from (1) because—and apparently only because—if (N) is false there will be “dependent deductive demonstrations.” Now a “deductive demonstration,” D , would consist of an argument where the premisses entail the conclusion such that the following are true:

- (a) All the propositions which are the premisses of D are cognitive.
- (b) The proposition which is the conclusion of D is ethical.

There are two ways in which an alleged deductive demonstration could be dependent. First, it will be dependent if the truth of (a) along with a specification of the premisses of D imply that the proposition which is the conclusion of the deductive demonstration is cognitive (where one does not need to rely on the validity of D and the inheritance of cognitivity to show this). Secondly, it will be dependent if the truth of (b) along with a specification of the conclusion of D imply that one or more of the propositions which are premisses of the deductive demonstration is ethical. In either of these cases the existence of a dependent deductive demonstration requires that there be a certain proposition which is both cognitive and ethical. Let me give some examples of what a *dependent* deductive demonstration would be like. If the entailment in an alleged deductive demonstration is of the form p entails p (where p is both ethical and cognitive), then this alleged deductive demonstration would be dependent. Whatever one does to show the premiss cognitive will show the conclusion cognitive and whatever is done to show the conclusion ethical will demonstrate that the premiss is ethical as well. This is just because the premiss is the conclusion. However, in other cases whether a deductive demonstration is dependent will be more problematic. Let the alleged deductive demonstration have this form:

- (i) C
- (ii) $C \supset E$
- (iii) E

where we will stipulate that E is ethical and that C is cognitive, but that neither of these claims implies the other. If this really is a deductive demonstration, then (ii) is cognitive. Now if the fact that (ii) is cognitive implies that E is cognitive, then the deductive demonstration here is dependent. On the

other hand, if it is the case—although this is unlikely—that any material implication is cognitive if its antecedent is and this is the *only* reason (ii) is cognitive, then the fact that (ii) is cognitive does not imply E is cognitive and the deductive demonstration will be independent.

Now it is important to note that claim, call it (C), that there are no dependent deductive demonstrations is equivalent to (N). First, (N) entails (C). If no proposition is both cognitive and ethical, then no dependent deductive demonstration exists since any such demonstration requires that such a proposition exist. Conversely, (C) entails (N) because $\sim(N)$ entails $\sim(C)$. If there existed proposition p which was both ethical and cognitive then this fact along with the trivial argument *if p then p* would constitute a dependent deductive demonstration. Thus, (C) would be false. So we may say that (C) is equivalent to (N). Now insofar as (1) rules out dependent deductive demonstrations it entails (C) and thus it must entail (N). But we should not suppose that because (1) entails (N) in this way it is ever likely to provide a basis for showing (N) to be true. It is more likely that (N) will be one of the things that has to be shown in order to show (1). In practice it appears we shall not be able to show—which is one thing (1) asserts—that there are absolutely no *dependent* deductive demonstrations except by first showing (N), i.e. that no proposition is both ethical and cognitive. So it would seem that little headway is made by replacing the issue in (N) with the issue in (1).

Still there is something about the *no-cught-from-is* claim which is not simply a matter of presupposing the truth of (N). Suppose we strip away from (1) that feature which seems to require our first knowing (N) to be true, *viz.* the ruling out of *dependent* deductive demonstrations. We are left with a more restricted, but more interesting, thesis than (1) which we may consider, *viz.*:

- 1' There is no independent deductive demonstration from cognitive premisses to an ethical conclusion.

Unlike (1), (1') does not entail (N). Although it may be part of the sceptic's claim, it is not the whole. Even if (1') were true, the following possibilities would distress any real sceptic: First, some ethical proposition might simply be a cognitive one even though it cannot be made the conclusion in an independent deductive demonstration. That Jones ought to return the book to Smith might be simply a fact about the situation not to be

confused with, or deduced from, other facts. Secondly, it might be thought that there is some independent inductive demonstration from cognitive premisses to an ethical conclusion where the premisses are thought to give some kind of inductive support to the *truth* of the conclusion. As might be expected, a *dependent* inductive demonstration is one in which the fact that the premisses are cognitive requires that the conclusion be cognitive as well, or else one in which the fact that the conclusion is ethical requires that at least one of the premisses be ethical. Again, there might be a kind of demonstration involving a third form of inference besides deduction and induction such that the *truth* of an ethical conclusion could be supported in some way by the truth of the premisses. As before, it would be independent demonstrations which we have in mind. Now if any one of these three is the case, (N) is not true. So it follows that the truth of (1') is not sufficient for (N) to be true. Thus understood, the *no-ought-from-is* dictum takes its place as a part of the non-cognitivist thesis—an important part, but not the whole. While entailed by non-cognitivism, it does not entail non-cognitivism.

3. THE NATURALISTIC FALLACY

Let us turn now to that other catchphrase of the sceptic, the so-called naturalistic (or definst) fallacy. That is, we must now consider how claim (2) fits in with (N) and (1'). First let us examine whether (2) entails (1'). We should note that Frankena's discussion might lead us to believe that (2) entails (1), but in fact there seems to be no reason to suppose it entails (1), or (1'), or (N). Let us suppose that (2) is true, i.e. that no ethical property can be properly defined in terms of descriptive ones. What prevents there being a simple proposition *p* which is both ethical and cognitive, so that both (1) and (N) would be false? My rhetorical question is not an argument but it does show that further conceptual connections have to be made, e.g. between descriptive characteristics and cognitive propositions. For example, we might want to say that any simple characteristic can be broken down to the form: $P(x_1, x_2, \dots, x_n)$, where *P* is an *n*-place predicate, and furthermore, that this *P* will be descriptive if and only if any proposition of the form: $P(a_1, a_2, \dots, a_n)$ is cognitive while *P* will be ethical if and only if any corresponding proposition is ethical. But there remain other

problems. There will be difficulties about complex propositions. Consider, for example, this argument:

- (i) aD
- (ii) $aD \supset aE$
- (iii) aE .

Where *E* is some ethical characteristic, *D* is some descriptive characteristic, and *a* is some individual. By hypothesis, (i) is cognitive and (iii) is ethical. The problem is about (ii). This premiss, it should be noted, is not a definition but only a complex proposition. As such it is consistent with the truth of the claim (2) which we continue to suppose is true. The problem is whether, under this supposition we must suppose (1') is also true. Of course, if the above deduction holds and if (ii) can be shown to be cognitive (in an independent demonstration), we must give up (1'). Now since I cannot prove otherwise, the sceptic can continue to maintain (1') under these conditions simply by maintaining that (ii) is not cognitive. This is his way out. Still, I think we should ask him whether he has any good reason for deciding to classify (ii) as non-cognitive except some blind devotion to sceptical theses such as (1'), (1), or (N). For what we would really like here is some sort of general criterion of a cognitive proposition which could be applied to complex propositions and which would have some independent plausibility. We are not impressed by criteria which have nothing to recommend them except that they allow the sceptic to continue to maintain his favorite theses. Now of course the observations so far in this paragraph do not amount to an argument that (2) does not imply (1'), (1), or (N). At best they have disposed the reader to be a bit more open-minded about the issues. But as a matter of fact Castañeda has given some rather convincing considerations for not believing that (2) entails (1) or (1').⁸ He points out that, although possibility cannot be defined in terms of reality, nevertheless *p* does entail *possibly p*. Likewise, he points out, we cannot define negation in terms of biconditionality. Nevertheless $p \equiv q$ implies $\sim p \equiv \sim q$.

On the other hand we might note that (1'), as well as (N), seems to entail (2). If (2) is false, so are (1') and (N). Suppose an ethical property (e.g. good) were definable in terms of a descriptive one (e.g. sought by all men). Then we could deduce:

- (i) pleasure is sought by all men
- (ii) what is sought by all men is good (definition)
- (iii) pleasure is good.

⁸ Hector-Neri Castañeda, *The Structure of Morality* (Springfield, Ill., 1974), p. 147 footnote 5.

Thus (1') would be false. Furthermore, (N) would be false assuming, as we are, that cognitivism is hereditary over deduction. There can be no real doubt that the conclusion is ethical, nor that the first premiss is cognitive. The second premiss, which would be necessarily true in virtue of definition, is as such to be classified as cognitive. For that matter, even a claim based on a false definition would appear to be cognitive. The sceptic might argue that the definition in question is really ethical, but whether this is so is irrelevant. We cannot simply assume that this prohibits it from being cognitive as well. That there might be necessary truths involving ethical terms should not, in principle, bother us for there surely are such in deontic logic. Now if someone should deny any of the above assertions about how to classify the premisses and conclusions of this argument, we would want to closely examine his reasons to make sure he is not taking us in circles. For example, we will not allow him to argue that the conclusion is not ethical *because* it follows from cognitive premisses or, again, that one of the premisses *must* be ethical since the conclusion is ethical. That kind of blind devotion to a thesis is not a philosophical virtue.

In summary, (N) entails (1) which entails its weaker form (1') which strictly implies (2). On the other hand (2) does not seem to entail any of these. This leads us to believe that (2) is a rather limited thesis, that (1') is more encompassing, and that (N) is the most encompassing thesis. And I think this is right. But as a matter of fact Frankena's discussion, when applied to these claims, would actually lead us to take a conflicting view, i.e. that (2) is, in certain respects at least, a more perspicacious formulation of scepticism than either (1') or (1).⁹ Frankena suggests that what the sceptic needs to attack is not the validity of the example argument, nor the cognitivism of its second premiss, but the truth of that premiss. Thus (2) and not (1') seems to be the important issue. However what this objection shows, I think, is not that we must accept (2) as a better formulation of scepticism in ethics but that (1') needs to be modified slightly. Claim (1') excludes the possibility that an ethical conclusion might be derived from cognitive premisses among which a *false* necessity claim is made (perhaps on the basis of a mistaken analysis of an

ethical term). Frankena's example is a case of this, it would appear. So the sceptic needs to modify (1') to:

- 1". There is no independent deductive demonstration from cognitive premisses (none of which make a false necessity claim) to an ethical conclusion.

where it may be observed that all the claims made for (1') until this present point in the discussion also holds for (1") as well.

Now the point to see here is that (1") is a more important and a more general sceptical claim than (2). Frankena's discussion ignores the way in which the issue of deduction, i.e. (1"), is already involved in the claim made by premiss (ii) of his example. To see this let us consider the possibility that we do not need this premiss to get a valid argument, that we can derive (iii) directly from (i). Frankena considers this possibility in his discussion but decides this shortened argument is really an enthymeme where (ii) is the suppressed premiss.¹⁰ Frankena says this, we may surmise, because he sees that without (ii) the argument is not *logically* valid. But logical validity is not exactly the same question as whether (i) entails (iii). It can be objected that what Frankena has been doing here is to restate the inference from the premiss to the conclusion as a further premiss.¹¹ Notice that (ii), unlike (i) and (iii), states a necessary truth. And as we learn from Lewis Carroll's "What the Tortoise said to Achilles" it is always possible to claim an argument needs a further (necessary) premiss. Perhaps this is not a knock-down argument. It is merely arbitrary, not inconsistent, to say we do not need any of the infinitely many extra premisses which Lewis Carroll's tortoise would require of us while maintaining that we do need the extra premiss in Frankena's example. Still, it is difficult to see what plausible argument there could be for excluding the former which would not be an argument for excluding the latter as well. If this observation is not convincing, the argument is knock-down for those who take entailment to be strict implication. If a premiss is a necessary truth, it will be true under all circumstances. Thus, if only the other premisses are true, the conclusion will be. But this is to say that these premisses alone strictly imply, or entail, the conclusion.

⁹ Frankena, *op. cit.*, p. 469.

¹⁰ Frankena, *op. cit.*, p. 468.

¹¹ Cf. Rynin, *op. cit.*, p. 315 and Bennett, *op. cit.*, pp. 230 ff. For a contrary view see the discussion of "suppression" in Richard and Valerie Routley, *op. cit.*

So we want to say that deleting (ii) does not affect the entailment. Of course at the same time it must be admitted that if (ii) is not a necessary truth (whether in virtue of a definition or in some other way) there is no entailment from (i) to (iii). The question of whether the extra premiss is a necessary truth is equivalent (but perhaps not identical) to the question of whether there is an entailment. If the second premiss of Frankena's argument is a necessary truth (perhaps because there is a definition or philosophical analysis of ethical characteristics in terms of descriptive ones), then we can deduce an *ought* from an *is* in an independent deductive demonstration. So $\sim(2)$ entails $\sim(1'')$, or $(1'')$ entails (2). Having said this, we can see that the sceptic's $(1'')$ should really be broken down to two claims, the second of which captures the concerns about avoiding "naturalistic fallacy" found in claim (2)—and more besides. Claim $(1'')$ breaks down to:

- A. There is no independent deductive demonstration from a set of non-necessary cognitive premisses to an ethical conclusion where the conclusion follows in virtue of the rules of *formal logic*.
- B. There is no independent deductive demonstration from a set of non-necessary cognitive premisses to an ethical conclusion where the conclusion follows in virtue of a *non-formal* entailment.

Where (B) is equivalent to:

- B'. There is no independent deductive demonstration from a set of cognitive premisses none of which makes a necessity claim which is false where the conclusion follows in virtue of the rules of *formal logic*

which may be more palatable to some philosophers, e.g. those who wish to reduce all entailments to rules of formal logic or those who feel that all the terms in the conclusion must be mentioned somewhere in the premisses,¹² or those who wish to maintain, in non-trivial fashion, that an ethical conclusion cannot follow from non-ethical (not, note, non-cognitive) premisses. With regard to (B), any alleged non-formal entailment ruled out will be equivalent to an alleged necessary truth ruled out by (B'). To rule out such a necessary truth one must not only attack those defenses which appeal

to definitions, but also those which would appeal to meanings of terms, synonyms, identity, philosophical intuition of necessary truth, conceptual analysis, or whatever. Anyone who holds (B) or (B') will, in particular, hold that ethical terms are not properly definable in terms of cognitive ones, i.e. he will hold (2). But this will only be a part of what he holds. There are many alternative grounds besides definition on which one might argue that there is a necessary truth which allows one to go from cognitive premisses to an ethical conclusion.

Thus (B) and (B') really give us a much broader understanding than (2) does of the supposed naturalistic fallacy. The naturalistic fallacy in this broader (and, admittedly, un-Moorean) sense would be the alleged mistake of supposing there is a necessary truth such that (B') is false, or a non-logical entailment such that (B) is false. If there is such a mistake, it would not be entirely misleading to call it a fallacy. For the dispute here is not about some ordinary matter of fact but about entailments and necessary truths. Of course there may be honest philosophical disagreement concerning what is a necessary truth or what is non-logically entailed. Furthermore, it is not an argument simply to label an opponent's view a "fallacy." Still, if (B) or its equivalent can be demonstrated, it is not inappropriate to use the term "fallacy." Whether it can be demonstrated is, of course, the question. It is also important to note that Moore's open question test—the main argument employed by non-cognitivists to demonstrate that the fallacy is a fallacy—can be used in regard to (B) and (B'). That is, if the open question argument works at all, it will not merely show (2) but also (B) and (B'). At the same time there is no more encompassing claim which has any chance of being demonstrated by this argument. Thus, I conclude, we may eliminate (2), the naturalistic fallacy thesis, in favour of (B), or its equivalent, which—as I have pointed out—is really a part of claim $(1')$, the no-*ought*-from-*is* thesis. While (B) is superior to (2), it is still a very limited claim. It needs to be supplemented by (A) to get $(1'')$. Even this falls short of (N), non-cognitivism. Still, (B) is an important claim for the non-cognitivist to make, even if it is not sufficient for non-cognitivism. It blocks off, as Frankena's example shows, one historically popular route to cognitivism.

Australian National University

Received May 4, 1976

¹² Frankena claims in his *Ethics*, 2nd ed. (Englewood Cliffs, N.J., 1973), p. 97 that we cannot deduce *A is C* from *A is B* without introducing some premiss connecting *B* and *C*. Similarly he argued in "The Naturalistic Fallacy," p. 468 that "it is invalid to argue that Croesus is rich because he is wealthy." For an adequate reply see Rynin, *op. cit.*, pp. 310–311.

VI. THE ARGUMENT FROM ANALOGY IS *NOT* AN ARGUMENT FOR OTHER MINDS

R. I. SIKORA

THE most common and for many the most forceful objection to the argument from analogy for other minds is that it must rely on a single case, namely, one's own. The argument is held to be weak because it must generalize from that one case to a vast number of other cases. A standard reply (used by Ayer and many others) is that we do not need to proceed from a single case of one mind correlated with one body, that we can proceed instead from an enormous number of correlations between instances of behavior and experiences. The experiences and the instances of behavior all belong to the same person, but this does not preclude there being an enormous number of correlations between the two sorts of event. I think that this response is basically correct although it has been thought to suffer from certain weaknesses which I shall consider in some detail.

Another objection to the argument from analogy is that whether or not it proceeds from a single case in the attempt to show that experiences are associated with other bodies, it *must* proceed from a single case in attempting to defend the more ambitious claim that at least one mind is associated with every other active human body. It is claimed that the only analogical argument that could support this conclusion is an analogical argument based on the fact that at least one mind has been associated with one's own body throughout its existence. For the argument from analogy to take us all the way to the conclusion that every other body has at least one mind, it would indeed have to rely in its final stage on a single case; but I will try in the first section to show that it isn't necessary to rely on the argument to establish that much. When properly understood, the argument from analogy is not in itself an argument for other minds.

I

I believe that once I can show by multiple-case analogical arguments that it is very highly probable that other bodies have experiences and that it is probable that at various times they are

having certain specific sorts of experiences (particularly ostensible memories), I can justify the hypothesis that at least one person is associated with every other body during its active existence (hypothesis PB) in much the same way as I can justify the similar hypothesis in regard to myself. And this is a method which does not involve any form of the argument from analogy and therefore *a fortiori* does not involve a single-case argument from analogy.

How then do I justify the hypothesis in regard to myself? At any given time, I have many memories of experiences which I seem to have had in the past. None of these memories associate these experiences with other bodies, and though some memories may not be associated with *any* body, I have other memories which make it reasonable for me to suppose that all of my experiences are causally dependent on my body in a way in which they are not dependent on the body of anyone else. I have had, of course, many more experiences than I can remember but there is no reason to suppose that the experiences I cannot remember differ from the ones that I can in a way that would justify me in thinking that the former were associated with some other body rather than with my own. On the basis of these considerations it is reasonable for me to infer that all of my experiences have been associated with the same body, and therefore that at least one person has animated my body throughout its existence.

When I turn to other bodies, I need only the suppositions (obtained, as I shall argue, through analogical arguments from a multiplicity of cases) that (i) those bodies have been associated with mental events throughout their lives, and (ii) that some of those mental events are ostensible memories that resemble my own, in order to defend a hypothesis for other bodies similar to that which I formed about myself. I am not offering here a single-case argument from analogy to the effect that since my body has been associated with a person, other bodies must be associated with persons. I am instead presenting a hypothesis in

regard to those other bodies—namely, hypothesis PB, and defending it in the same manner as I have defended the similar hypothesis about myself—except that in the case of other bodies I must ascertain first by analogical arguments relying on a number of cases that those bodies are associated with mental events, some of which are ostensible memories similar to my own. I shall call this defence of hypothesis PB the memory-of-same-body defence.

To make it doubly clear that the memory-of-same-body defence of hypothesis PB is not a single-case argument from analogy, consider a case in which a person would not be in a position to support hypothesis PB by a single-case analogical argument yet could still support it by the memory-of-same-body defence. A is suffering from amnesia; he can remember no events prior to the last six months. He is therefore not in a position to determine that his own experiences have always been associated with the same body: for all he knows, they might have been associated with another body before the onset of his amnesia. Thus he can't defend hypothesis PB by a single-case argument based on himself. He can, however, still defend that hypothesis by the memory-of-same-body defence. After using a multiple-case argument from analogy to determine that other persons have ostensible memories, he can use their memory reports to justify the claim that whether or not he himself has always been associated with the same body, there are good grounds for supposing that every other body has had at least one person associated with it throughout its existence. (The same would hold true in the far-fetched case where A has memories according to which his experiences were associated with one body during the first half of his life and with another during the second half. In this case, an argument from analogy would in fact suggest that hypothesis PB was false, but the claim that it was true would still be supported by the memory-of-same-body defence.)¹

have used this sort of multiple-case argument from analogy to justify the claim that ostensible memories similar to my own are associated with other bodies, I don't need an additional argument from analogy (and, *a fortiori*, not a single-case argument from analogy) to justify the claim that at least one person is associated with each human body. But it has been claimed that multiple-case arguments from analogy can't justify the contention about ostensible memories or indeed justify any claim that experiences are associated with other bodies. It is commonly supposed that although one can correlate vast numbers of one's own experiences with instances of one's own behavior, all correlations after the first are redundant so that one must in effect rely on an argument from a single case and that such a basis is unsatisfactory. Budlong argues (*Analysis* 35.3) that just as it would be redundant for a person testing a bottle of milk for freshness to take a second glass to serve as a supplement to the first, it would also be redundant to use more than one correlation between your own mental state and your behavior. But let's suppose that you really tried to make the argument from analogy work on the basis of a single case. You would have to imagine that you had forgotten all of your previous experiences and that you had observed only a single correlation of an instance of behavior with a mental event. Perhaps you would be entitled to infer that there were associated mental events when other bodies exhibited behavior that was just like your behavior in the case you had observed, but any inference that went beyond this either about your own body or about the bodies of others would be on a very shaky foundation indeed. I grant that you don't need the vast multitude of correlations that one could find by a careful and thorough use of memory, or by, for example, systematically recording your experiences and the accompanying behavior, but you obviously need much more than a single correlation.

II

If my preceding contentions are correct, once I

III

If I am right, multiple-case arguments from

¹ It has been suggested to me that although multiple-case arguments from analogy aren't suitable to the defence of the claim that there are other minds if that claim is construed traditionally as the claim that for every other human body there is one mind or at least one mind, it can be used to show that there are other minds (leaving aside the question of whether they are associated with just one body). But even in this case a multiple-case argument from analogy wouldn't take us far enough. After it was determined by a multiple-case argument that various sorts of experiences are associated with other bodies, another sort of argument is needed to show that these experiences are joined together into minds. If it were claimed that a single-case argument from analogy is needed for this purpose, I can remind my opponent that I have already shown that such a single-case argument is not needed to support the conclusion that there are other minds associated with a single body, and therefore, *a fortiori*, is not needed to show that there are other minds.

analogy are not redundant, but perhaps even with a multiplicity of correlations the argument won't work. I shall consider a number of attempts to show that this is the case.

1. It has been urged that even if it were granted that a number of correlations between my mental states and my behavior are better than just one, the class of actions where I have observed this correlation is only a minute fraction of the class of actions about which I want to make an inference. My sample class is far too small to allow me to have anything like the confidence that we are entitled to have about the existence of mental states in other persons.

But as a matter of fact we make confident judgments about the behavior of sub-atomic particles on the basis of observations of a much smaller fraction of sub-atomic particles. Further, most opponents of the argument from analogy would remain dissatisfied even if we were only making an inference about *one* other person; or for that matter, even if we were only inferring that a single action of another person was accompanied by a single experience.

2. It could be argued that the behavior of other people differs too much from one's own behavior for the latter to serve as the basis for inference. But even if everyone had exactly the same sort of body and behaved in exactly the same way, opponents of the argument from analogy would remain dissatisfied. And no one suggests that the probability of the claim that other persons have experiences varies to any significant extent depending on the similarity of those other persons to oneself. The objection seems to be rather that my sample does not include anyone who is not myself. When it is reduced to this, it carries little weight. It seems more important than it is simply because it is logically impossible to verify the claim that this particular difference is, as it would seem to be, unimportant to the analogy. If this is the true source of the objection, it comes down to no more than the objection that the conclusion of the argument is unverifiable, an objection that Hyslop and Jackson² have shown to be ineffectual.

3. It has been argued³ that (i) since I only observe that most of my manifestations of pain-behavior are accompanied by pain, I can only infer that (ii) many instances of behavioral states

are associated with mental states and that it does not follow from this that (iii) the common sense belief that *every* body which behaves in a manner similar to my own will have many instances of mental states associated with its behavioral states is correct.

But despite the fact that it is only in most cases (but not in all) that my own pain behavior is accompanied by pain, it is incorrect to infer from this that my observations entitle me only to conclude that many rather than all instances of behavioral states are associated with mental states. On those occasions when I engage in pain behavior but am free from pain, my pain behavior is nonetheless associated with *some* other mental state. Thus one is entitled to infer that, for *every* body which behaves much as one's own, that many instances of its behavioral states are associated with mental states.

So far so good, but we expect the argument from analogy to yield not merely the conclusion that all other human bodies are associated with mental events but also that they are associated at some time in their histories with certain sorts of mental event—with, for example, pains and ostensible memories (henceforth memories). Now as I observed before, I do not always feel pain when I exhibit pain behavior. Could it be argued on this basis that although the argument from analogy entitles me to suppose that every human body has associated mental events, it does not entitle me to suppose that *every* human body at some time or other has pains associated with it? Consider the argument just criticized with the modification that we stick to pains throughout rather than switching from pains to mental events halfway through.

1. Usually I feel pain when I exhibit pain-behavior.
2. Therefore many instances of pain behavior are associated with pain.
3. It is nonetheless possible that when certain human bodies exhibit pain behavior, this pain behavior is never accompanied with pain.
4. Conclusion! A multiple-case argument from analogy fails to show that it is certain that all human bodies are sometimes accompanied with pain.

The trouble with this is that its conclusion isn't particularly damaging to multiple-case arguments

² A. Hyslop and F. C. Jackson, "The Analogical Inference to Other Minds," *American Philosophical Quarterly*, vol. 9 (1972), pp. 168-176.

³ *Ibid.*, pp. 168-173.

from analogy. It can be admitted that we can't be certain that for every single human body some instances of its pain-behavior are associated with pain, so long as the possibility is sufficiently slight. It might be objected that the plain man wouldn't admit that there is even an extremely slight possibility of such exceptions. However, a philosopher needn't be unduly disturbed if this turns out to be another case where we can only hope for moral certainty, even if the plain man thinks (or is supposed to think) that we can have absolute certainty.

But perhaps the argument can be revised so as to provide a sufficiently strong conclusion to create a real problem. Consider the following:

I observe a number of correlations between my own pain and my own pain behavior. However, I also note that there are a small but significant number of occasions when I exhibit pain behavior but feel no pain. Accordingly, whenever I observe pain behavior in another person, an analogy with myself would lead me to conclude that there is a small but significant chance that he is not in pain. It follows from this that if I were to observe all the pain behavior of everyone who exists, statistical considerations would make it fairly probable that there are some human bodies such that no instance of their pain behavior is accompanied by pain. Therefore I am not entitled by a multiple-case argument from analogy to be as sure as I am in fact entitled to be that every other body is at some time associated with pain.

This may look plausible but it isn't; the multiple-case argument from analogy can be buttressed with too much additional evidence for the conclusion to survive. In the first place, I could note that there are only certain circumstances in which I exhibit pain behavior when I am not in pain, and if I could observe other persons long enough (which an objector to the argument from analogy would obviously allow) I could surely find occasions for any other person when he exhibited pain behavior and these circumstances didn't pertain. More important, I needn't limit myself to correlations between pain and pain behavior. I can also consider correlations between pain and pain stimuli because in the case of certain kinds of pain stimuli, there is a far greater correlation between my being subjected to the stimuli and feeling pain than there is between my exhibiting pain behavior and feeling pain. Finally, if one went on to talk about neural responses to stimuli, a real doubter could be led to find correlations between his own neural states and feelings of pain which would presumably be even

tighter than the correlations between pains and pain-stimuli. Any doubts that remained after this would I think be perfectly tolerable.

IV

I have argued that we can establish with a reasonable degree of certainty that experiences including ostensible memories are associated with other bodies and that once this is granted, we don't need an additional argument from analogy—this time a single-case argument—to show that there is at least one mind associated with every body. It might still be objected that it is unnecessary to use the argument from analogy because we have in any case a better reason for supposing that other bodies are 'inhabited' by minds: we are justified in hypothesizing the existence of other minds in order to provide a causal explanation for the rational behavior of other bodies. But this explanation would be false if epiphenomenalism were true. Probably epiphenomenalism is false but the probability of that does not appear to be as great as the probability that there are other minds or the probability afforded by the argument from analogy. Therefore it is unsatisfactory to rest the claim that there are other minds exclusively or even primarily on any argument that relies on the falsity of epiphenomenalism.

It has been suggested in response to my last remarks that there is no reason at all to give any weight at all to any analogical argument unless it involves the premiss that certain events *cause* other events; and, specifically, that an analogical argument for the existence of 'other' mental events absolutely requires the premiss that some of one's own behavior is caused by associated mental events; and thus of course to require the falsity of epiphenomenalism. I agree to the extent of granting that the argument must claim that there are causal relations between mental events and physical events; but in terms of the argument *it doesn't matter which way the causal relations go*. Interactionism would be fine, but from the point of view of the argument it would be equally satisfactory if epiphenomenalism were true for it too would guarantee that mental events accompany certain sorts of physical events; and it would also for that matter be satisfactory if mental events caused physical events but physical events never caused mental ones. I should add that the question of whether the truth of epiphenomenalism is compatible with the success of the argument from

analogy is of more than academic interest: Shaffer, certainly one of the most sophisticated writers on the philosophy of mind, thinks that there is a considerable chance that epiphenomenalism is true.⁴

Alternatively, it can be argued that we are justified in supposing that there are other minds because the hypothesis that all living human bodies are accompanied by minds is in some sense simpler than the hypothesis that you are an exception to an otherwise universal rule that such bodies are not accompanied by minds. The hypothesis that there are other minds does not claim that mental events cause physical events or even that some psychophysical events cause others so its truth could be more probable than the claim that epiphenomenalism is false.

However, in the last analysis it relies itself on an argument from analogy that would seem to be weaker than the one we rely on in the argument from analogy itself. When confronted with two hypotheses both of which account for the facts, the best reason for accepting the simpler one as correct is the fact that in the past further investigations have confirmed the simpler of two rival hypotheses more often than the more complex one. But the simplicity of the hypothesis that there are other minds is not just like the simplicity of the various other simple hypotheses that have been confirmed in the past so that we are relying here on no more than an analogy between the simplicity of the hypothesis that there are other minds and that of various other hypotheses. Further, I feel more

confidence in my intuition that its being other persons is irrelevant to the analogy in the case of the argument from analogy than in my intuition that the difference between the simplicity of the hypothesis that there are other minds and the simplicity of the various other hypotheses is irrelevant. Accordingly, here again it seems to me that the argument from analogy provides the best support for our belief that experiences are associated with other bodies.

SUMMARY

The argument from analogy should not be construed as an argument for other minds; it is rather an argument that certain sorts of experience are associated with other bodies. It is a multiple-case argument. The multiplicity of cases do not make it redundant and the number of cases is adequate to establish the conclusion of the argument with a reasonable degree of certainty. Once it has been established that ostensible memories similar to my own are associated with other bodies, I am entitled to infer without being forced to use an additional argument from analogy (this one proceeding from a single case) that at least one person is associated with every other active human body. The reports of ostensible memories that I am given would entitle me to make this hypothesis about other bodies even if my own ostensible memories did not entitle me to make it about myself.⁵

University of British Columbia

Received March 3, 1976

⁴ Cf. *Philosophy of Mind* (Englewood Cliffs, 1968), p. 76.

⁵ I am indebted to Jonathan Bennett for comments and criticisms.

VII. THE PRAGMATICS OF EXPLANATION

BAS C. VAN FRAASSEN

THERE are two problems about scientific explanation. The first is to describe it: when is something explained? The second is to show why (or in what sense) explanation is a virtue. Presumably we have no explanation unless we have a good theory; one which is independently worthy of acceptance. But what virtue is there in explanation over and above this? I believe that philosophical concern with the first problem has been led thoroughly astray by mistaken views on the second.

I. FALSE IDEALS

To begin I wish to dispute three ideas about explanation that seem to have a subliminal influence on the discussion. The first is that explanation is a relation simply between a theory or hypothesis and the phenomena or facts, just like truth for example. The second is that explanatory power cannot be logically separated from certain other virtues of a theory, notably truth or acceptability. And the third is that explanation is the overriding virtue, the end of scientific inquiry.

When is something explained? As a foil to the above three ideas, let me propose the simple answer: *when we have a theory which explains*. Note first that "have" is not "have on the books"; I cannot claim to have such a theory without implying that this theory is acceptable all told. Note also that both "have" and "explains" are tensed; and that I have allowed that we can have a theory which does not explain, or "have on the books" an unacceptable one that does. Newton's theory explained the tides but not the advance in the perihelion of mercury; we used to have an acceptable theory, provided by Newton, which bore (or bears timelessly?) the explanation relationship to some facts but not to all. My answer also implies

that we can intelligibly say that the theory explains, and not merely that people can explain by means of the theory. But this consequence is not very restrictive, because the former could be an ellipsis for the latter.

There are questions of usage here. I am happy to report that the history of science allows systematic use of both idioms. In Huygens and Young the typical phrasing seemed to be that phenomenon may be explained *by means of* principles, laws and hypotheses, or *according to* a view.¹ On the other hand, Fresnel writes to Arago in 1815 "Tous ces phénomènes . . . sont réunis et expliqués par la même théorie des vibrations," and Lavoisier says that the oxygen hypothesis he proposes *explains* the phenomena of combustion.² Darwin also speaks in the latter idiom: "In scientific investigations it is permitted to invent any hypothesis, and if it explains various large and independent classes of facts it rises to the rank of a well-grounded theory"; though elsewhere he says that the facts of geographical distribution are *explicable on* the theory of migration.³

My answer did separate acceptance of the theory from its explanatory power. Of course, the second can be a reason for the first; but *that* requires their separation. Various philosophers have held that explanation logically requires true (or acceptable) theories as premises. Otherwise, they hold, we can at most mistakenly believe that we have an explanation.

This is also a question of usage, and again usage is quite clear. Lavoisier said of the phlogiston hypothesis that it is too vague and consequently "s'adapte a toutes les explications dans lesquelles on veut le faire entrer."⁴ Darwin explicitly allows explanations by false theories when he says "It can hardly be supposed that a false theory would

¹ I owe these and following references to my student Mr. Paul Thagard. For instance see C. Huygens, *Treatise on Light*, tr. by S. P. Thompson (New York, 1962), pp. 19, 20, 22, 63; Thomas Young, *Miscellaneous Works*, ed. by George Peacock (London, 1855), Vol. I, pp. 168, 170.

² Augustin Fresnel, *Oeuvres Complètes* (Paris, 1866), Vol. I, p. 36 (see also pp. 254, 355); Antoine Lavoisier, *Oeuvres* (Paris, 1862), Vol. II, p. 233.

³ Charles Darwin, *The Variation of Animals and Plants* (London, 1868), Vol. I, p. 9; *On the Origin of the Species* (Facs. of first edition, Cambridge, Mass., 1964), p. 408.

⁴ Antoine Lavoisier, *op. cit.*, p. 640.

explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified."⁵ More recently, Gilbert Harman has argued similarly: that a theory explains certain phenomena is part of the evidence that leads us to accept it. But that means that the explanation-relation is visible beforehand. Finally, we criticize theories selectively: a discussion of celestial mechanics around the turn of the century would surely contain the assertion that Newton's theory does explain many planetary phenomena, though not the advance in the perihelion of Mercury.

There is a third false ideal, which I consider worst: that explanation is the *summum bonum* and exact aim of science. A virtue could be overriding in one of two ways. The first is that it is a minimal criterion of acceptability. Such is consistency with the facts in the domain of application (though not necessarily with all data, if these are dubitable!). Explanation is not like that, or else a theory would not be acceptable at all unless it explained all facts in its domain. The second way in which a virtue may be overriding is that of being required when it can be had. This would mean that if two theories pass other tests (empirical adequacy, simplicity) equally well, then the one which explains more must be accepted. As I have argued elsewhere,⁶ and as we shall see in connection with Salmon's views below, a precise formulation of this demand requires hidden variables for indeterministic theories. But of course, hidden variables are rejected in scientific practice as so much "metaphysical baggage" when they make no difference in empirical predictions.

II. A BIASED HISTORY

I will outline the attempts to characterize explanation of the past three decades, with no pretense of objectivity. On the contrary, the selection is meant to illustrate the diagnosis, and point to the solution, of the next section.

1. Hempel

In 1966, Hempel summarized his views by listing two main criteria for explanation. The first is the criterion of *explanatory relevance*: "the explanatory information adduced affords good grounds for

believing that the phenomenon to be explained did, or does, indeed occur."⁷ That information has two components, one supplied by the scientific theory, the other consisting of auxiliary factual information. The relationship of providing good grounds is explicated as (a) implying (D-N case), or (b) conferring a high probability (I-S case), which is not lowered by the addition of other (available) evidence.

As Hempel points out, this criterion is not a sufficient condition for explanation: the red shift gives us good grounds for believing that distant galaxies are receding from us, but does not explain why they do. The classic case is the *barometer example*: the storm will come exactly if the barometers fall, which they do exactly if the atmospheric conditions are of the correct sort; yet only the last factor explains. Nor is the criterion a necessary condition; for this the classic case is the *paresis example*. We explain why the mayor, alone among the townsfolk, contracted paresis by his history of latent, contracted syphilis; yet such histories are followed by paresis in only a small percentage of cases.

The second criterion is the requirement of *testability*; but since all serious candidates for the role of scientific theory meet this, it cannot help to remove the noted defects.

2. Beckner, Putnam, and Salmon

The criterion of explanatory relevance was revised in one direction, informally by Beckner and Putnam and precisely by Salmon. Morton Beckner, in his discussion of evolution theory, pointed out that this often explains a phenomenon only by showing how it could have happened, given certain possible conditions.⁸ Evolutionists do this by constructing models of processes which utilize only genetic and natural selection mechanisms, in which the outcome agrees with the actual phenomenon. Parallel conclusions were drawn by Hilary Putnam about the way in which celestial phenomena are explained by Newton's theory of gravity: celestial motions could indeed be as they are, given a certain possible (though not, known) distribution of masses in the universe.⁹

We may take the paresis example to be explained similarly. Mere consistency with the theory is of

⁵ *Origin* (sixth ed., New York, 1962), p. 476.

⁶ "Wilfrid Sellars on Scientific Realism," *Dialogue*, vol. 14 (1975), pp. 606-616.

⁷ C. G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, New Jersey, 1966), p. 48.

⁸ *The Biological Way of Thought* (Berkeley, 1968), p. 176; this was first published in 1959.

⁹ In a paper of which a summary is found in Frederick Suppe (ed.), *The Structure of Scientific Theories* (Urbana, Ill., 1974).

course much too weak, since that is implied by logical irrelevance. Hence Wesley Salmon made this precise as follows: to explain is to exhibit (the) statistically relevant factors.¹⁰ (I shall leave till later the qualifications about "screening off.") Since this sort of explication discards the talk about modelling and mechanisms of Beckner and Putnam, it may not capture enough. And indeed, I am not satisfied with Salmon's arguments that his criterion provides a sufficient condition. He gives the example of an equal mixture of Uranium 238 atoms and Polonium 214 atoms, which makes the Geiger counter click in interval $(t, t+m)$. This means that one of the atoms disintegrated. Why did it? The correct answer will be: because it was a Uranium 238 atom, if that is so—although the probability of its disintegration is much higher relative to the previous knowledge that the atom belonged to the described mixture.¹¹ The problem with this argument is that, on Salmon's criterion, we can explain not only why there was a disintegration, but also why *that* atom disintegrated *just then*. And surely that is exactly one of those facts which atomic physics leaves unexplained?

But there is a more serious general criticism. Whatever the phenomenon is, we can amass the statistically relevant factors, as long as the theory does not rule out the phenomenon altogether. "What more could one ask of an explanation?" Salmon inquires.¹² But in that case, as soon as we have an empirically adequate theory, we have an explanation of every fact in its domain. We may claim an explanation as soon as we have shown that the phenomenon can be embedded in some model allowed by the theory—that is, does not throw doubt on the theory's empirical adequacy.¹³ But surely that is too sanguine?

3. Global Properties

Explanatory power cannot be identified with empirical adequacy; but it may still reside in the performance of the theory as a whole. This view is accompanied by the conviction that science does not explain individual facts but general regularities

and was developed in different ways by Michael Friedman and James Greeno. Friedman says explicitly that in his view, "the kind of understanding provided by science is global rather than local" and consists in the simplification and unification imposed on our world picture.¹⁴ That S_1 explains S_2 is a conjunction of two facts: S_1 implies S_2 relative to our background knowledge (and/or belief) K , and S_1 unifies and simplifies the set of its consequences relative to K . Friedman will no doubt wish to weaken the first condition in view of Salmon's work.

The precise explication Friedman gives of the second condition does not work, and is not likely to have a near variant that does.¹⁵ But here we may look at Greeno's proposal.¹⁶ His abstract and closing statement subscribe to the same general view as Friedman. But he takes as his model of a theory one which specifies a single probability space Q as the correct one, plus two partitions (or random variables) of which one is designated *explanandum* and the other *explanans*. An example: sociology cannot explain why Albert, who lives in San Francisco and whose father has a high income, steals a car. Nor is it meant to. But it does explain delinquency in terms of such other factors as residence and parental income. The degree of explanatory power is measured by an ingeniously devised quantity which measures the information I the theory provides of the explanandum variable M on the basis of explanans S . This measure takes its maximum value if all conditional probabilities $P(M_i/S_j)$ are zero or one (D-N case), and its minimum value zero if S and M are statistically independent.

Unfortunately, this way of measuring the unification imposed on our data abandons Friedman's insight that scientific understanding cannot be identified as a function of grounds for rational expectation. For if we let S and M describe the behavior of the barometer and coming storms, with $P(\text{barometer falls}) = P(\text{storm comes}) = 0.2$, $P(\text{storm comes}/\text{barometer falls}) = 1$, and $P(\text{storm comes}/\text{barometer does not fall}) = 0$, then the quantity I

¹⁰ "Statistical Explanation," pp. 173–231 in R. G. Colodny (ed.), *The Nature and Function of Scientific Theories* (Pittsburgh, 1970); reprinted also in Salmon's book cited below.

¹¹ *Ibid.*, pp. 207–209. Nancy Cartwright has further, unpublished, counter-examples to the necessity and sufficiency of Salmon's criterion.

¹² *Ibid.*, p. 222.

¹³ These concepts are discussed in my "To Save the Phenomena," *The Journal of Philosophy*, vol. 73 (1976), forthcoming.

¹⁴ "Explanation and Scientific Understanding," *The Journal of Philosophy*, vol. 71 (1974), pp. 5–19.

¹⁵ See Philip Kitcher, "Explanation, Conjunction, and Unification," *The Journal of Philosophy*, vol. 73 (1976), pp. 207–212.

¹⁶ "Explanation and Information," pp. 89–103 in Wesley Salmon (ed.), *Statistical Explanation and Statistical Relevance* (Pittsburgh, 1971). This paper was originally published with a different title in *Philosophy of Science*, vol. 37 (1970), pp. 279–293.

takes its maximum value. Indeed, it does so whether we designate M or S as explanans.

It would seem that such asymmetries as exhibited by the red shift and barometer examples must necessarily remain recalcitrant for any attempt to strengthen Hempel's or Salmon's criteria by global restraints on theories alone.

4. *The Major Difficulties*

There are two main difficulties, illustrated by the old paresis and barometer examples, which none of the examined positions can handle. The first is that there are cases, clearly in a theory's domain, where the request for explanation is nevertheless rejected. We can explain why John, rather than his brothers contracted paresis, for he had syphilis; but not why he, among all those syphilitics, got paresis. Medical science is incomplete, and hopes to find the answer some day. But the example of the uranium atom disintegrating just then rather than later, is formally similar and we believe the theory to be complete. We also reject such questions as the Aristotelians asked the Galileans: why does a body free of impressed forces retain its velocity? The importance of this sort of case, and its pervasive character, has been repeatedly discussed by Adolf Grünbaum.

The second difficulty is the asymmetry revealed by the barometer: even if the theory implies that one condition obtains when and only when another does, it may be that it explains the one in terms of the other and not vice versa. An example which combines both the first and second difficulty is this: according to atomic physics, each chemical element has a characteristic atomic structure and a characteristic spectrum (of light emitted upon excitation). Yet the spectrum is explained by the atomic structure, and the question why a substance has that structure does not arise at all (except in the trivial sense that the questioner may need to have the terms explained to him).

5. *Causality*

Why are there no longer any Tasmanian natives? Well, they were a nuisance, so the white settlers just kept shooting them till there were none left. The request was not for population statistics, but for the story; though in some truncated way, the statistics "tell" the story.

In a later paper Salmon gives a primary place to causal mechanisms in explanation.¹⁷ Events are

bound into causal chains by two relations: spatio-temporal continuity and statistical relevance. Explanation requires the exhibition of such chains. Salmon's point of departure is Reichenbach's *principle of the common cause*: every relation of statistical relevance ought to be explained by one of causal relevance. This means that a correlation of simultaneous values must be explained by a prior common cause. Salmon gives two statistical conditions that must be met by a common cause C of events A and B :

$$(a) P(A \& B/C) = P(A/C)P(B/C)$$

$$(b) P(A/B \& C) = P(A/C) \quad \text{"C screens off B from A."}$$

If $P(B/C) \neq 0$ these are equivalent, and symmetric in A and B .

Suppose that explanation is typically the demand for a common cause. Then we still have the problem: when does this arise? Atmospheric conditions explain the correlation between barometer and storm, say; but are still prior causes required to explain the correlation between atmospheric conditions and falling barometers?

In the quantum domain, Salmon says, causality is violated because "causal influence is not transmitted with spatio-temporal continuity." But the situation is worse. To assume Reichenbach's principle to be satisfiable, continuity aside, is to rule out all genuinely indeterministic theories. As example, let a theory say that C is invariably followed by one of the incompatible events A , B , or D , each with probability $1/3$. Let us suppose the theory complete, and its probabilities irreducible, with C the complete specification of state. Then we will find a correlation for which only C could be the common cause, but it is not. Assuming that A , B , D are always preceded by C and that they have low but equal prior probabilities, there is a statistical correlation between $\phi = (A \text{ or } D)$ and $\psi = (B \text{ or } D)$, for $P(\phi/\psi) = P(\psi/\phi) = 1/2 \neq P(\phi)$. But C , the only available candidate, does not screen off ϕ from ψ : $P(\phi/C \& \psi) = P(\phi/\psi) = 1/2 \neq P(\phi/C)$ which is $2/3$. Although this may sound complicated, the construction is so general that almost any irreducibly probabilistic situation will give a similar example. Thus Reichenbach's *principle of the common cause* is in fact a demand for hidden variables.

Yet we retain the feeling that Salmon has given an essential clue to the asymmetries of explanation. For surely the crucial point about the barometer is

¹⁷ "Theoretical Explanation," pp. 118-145 in Stephan Körner (ed.), *Explanation* (Oxford, 1975).

that the atmospheric conditions screen off the barometer fall from the storm? The general point that the asymmetries are totally bound up with causality was argued in a provocative article by B. A. Brody.¹⁸ Aristotle certainly discussed examples of asymmetries: the planets do not twinkle because they are near, yet they are near if and only if they do not twinkle (*Posterior Analytics*, I, 13). Not all explanations are causal, says Brody, but the others use a second Aristotelian notion, that of essence. The spectrum angle is a clear case: sodium has that spectrum because it has this atomic structure, which is its essence.

Brody's account has the further advantage that he can say when questions do not arise: other properties are explained in terms of essence, but the request for an explanation of the essence does not arise. However, I do not see how he would distinguish between the questions why the uranium atom disintegrated and why it disintegrated just then. In addition there is the problem that modern science is not formulated in terms of causes and essences, and it seems doubtful that these concepts can be redefined in terms which do occur there.

6. *Why-Questions*

A why-question is a request for explanation. Sylvain Bromberger called *P* the *presupposition* of the question *Why-P?* and restated the problem of explanation as that of giving the conditions under which proposition *Q* is a correct answer to a why-question with presupposition *P*.¹⁹ However, Bengt Hansson has pointed out that "Why was it John who ate the apple?" and "Why was it the apple which John ate?": are different why-questions, although the comprised proposition is the same.²⁰ The difference can be indicated by such phrasing, or by emphasis ("Why did *John* . . .?") or by an auxiliary clause ("Why did John rather than . . .?"). Hansson says that an explanation is requested, not of a proposition or fact, but of an *aspect* of a proposition.

As is at least suggested by Hansson, we can cover all these cases by saying that we wish an explanation of why *P* is true in contrast to other members of a set *X* or propositions. This explains the tension in our reaction to the paresis-example. The question why the mayor, in contrast to other townfolk generally, contracted paresis *has* a true

correct answer: because of his latent syphilis. But the question why he did in contrast to the other syphilitics in his country club, has no true correct answer. Intuitively we may say: *Q* is a correct answer to *Why P in contrast to X?* only if *Q* gives reasons to expect that *P*, in contrast to the other members of *X*. Hansson's proposal for a precise criterion is: the probability of *P* given *Q* is higher than the average of the probabilities of *R* given *Q*, for members *R* of *X*.

Hansson points out that the set *X* of alternatives is often left tacit; the two questions about paresis might well be expressed by the same sentence in different contexts. The important point is that explanations are not requested of propositions, and consequently a distinction can be drawn between answered and rejected requests in a clear way. However, Hansson makes *Q* a correct answer to *Why P in contrast to X?* when *Q* is statistically irrelevant, when *P* is already more likely than the rest; or when *Q* implies *P* but not the others. I do not see how he can handle the barometer (or red shift, or spectrum) asymmetries. On his precise criterion, that the barometer fell is a correct answer to why it will storm as opposed to be calm. The difficulty is very deep: if *P* and *R* are necessarily equivalent, according to our accepted theories, how can *Why P in contrast to X?* be distinguished from *Why R in contrast to X?*

III. THE SOLUTION

1. *Prejudices*

Two convictions have prejudiced the discussion of explanation, one methodological and one substantive.

The first is that a philosophical account must aim to produce necessary and sufficient conditions for theory *T* explaining phenomenon *E*. A similar prejudice plagued the discussion of counter-factuals for twenty years, requiring the exact conditions under which, if *A* were the case, *B* would be. Stalnaker's liberating insight was that these conditions are largely determined by context and speaker's interest. This brings the central question to light: what *form* can these conditions take?

The second conviction is that explanatory power is a virtue of theories by themselves, or of their relation to the world, like simplicity, predictive

¹⁸ "Towards an Aristotelian Theory of Scientific Explanation," *Philosophy of Science*, vol. 39 (1972), pp. 20-31.

¹⁹ "Why-Questions," pp. 86-108 in R. G. Colodny (ed.), *Mind and Cosmos* (Pittsburgh, 1966).

²⁰ "Explanations—Of What?" (mimeographed: Stanford University, 1974).

strength, truth, empirical adequacy. There is again an analogy with counterfactuals: it used to be thought that science contains, or directly implies, counterfactuals. In all but limiting cases, however, the proposition expressed is highly context-dependent, and the implication is there at most relative to the determining contextual factors, such as speakers' interest.

2. *Diagnosis*

The earlier accounts lead us to the format: C explains E relative to theory T exactly if (a) T has certain global virtues, and (b) T implies a certain proposition $\phi(C, E)$ expressible in the language of logic and probability theory. Different accounts directed themselves to the specification of what should go into (a) and (b). We may add, following Beckner and Putnam, that T explains E exactly if there is a proposition C consistent with T (and presumably, background beliefs) such that C explains E relative to T .

The significant modifications were proposed by Hansson and Brody. The former pointed out that the explanandum E cannot be reified as a proposition: we request the explanation of something F in contrast to its alternatives X (the latter generally tacitly specified by context). This modification is absolutely necessary to handle some of our puzzles. It requires that in (b) above we replace " $\phi(C, E)$ " by the formula form " $\psi(C, F, X)$." But the problem of asymmetries remains recalcitrant, because if T implies the necessary equivalence of F and F' (say, atomic structure and characteristic spectrum), then T will also imply $\psi(C, F', X)$ if and only if it implies $\psi(C, F, X)$.

The only account we have seen which grapples at all successfully with this, is Brody's. For Brody points out that even properties which we believe to be constantly conjoined in all possible circumstances, can be divided into essences and accidents, or related as cause and effect. In this sense, the asymmetries were no problem for Aristotle.

3. *The logical problem*

We have now seen exactly what logical problem is posed by the asymmetries. To put it in current terms: how can we distinguish propositions which are true in exactly the same possible worlds?

There are several known approaches that use impossible worlds. David Lewis, in his discussion of

causality, suggests that we should look not only to the worlds theory T allows as possible, but also to those it rules out as impossible, and speaks of counterfactuals which are counterlegal. Relevant logic and entailment draw distinctions between logically equivalent sentences and their semantics devised by Routley and Meyer use both inconsistent and incomplete worlds. I believe such approaches to be totally inappropriate for the problem of explanation, for when we look at actual explanations of phenomena by theories, we do not see any detours through circumstances or events ruled out as impossible by the theory.

A further approach, developed by Rolf Schock, Romane Clark, and myself distinguishes sentences by the facts that make them true. The idea is simple. That it rains, that it does not rain, that it snows, and that it does not snow, are four distinct facts. The disjunction that it rains or does not rain is made true equally by the first and second, and not by the third or fourth, which distinguishes it from the logically equivalent disjunction that it snows or does not snow.²¹ The distinction remains even if there is also a fact of its raining or not raining, distinct or identical with that of its snowing or not snowing.

This approach can work for the asymmetries of explanation. Such asymmetries are possible because, for example, the distinct facts that light is emitted with wavelengths λ , μ , . . . conjointly make up the characteristic spectrum, while quite different facts conjoin to make up the atomic structure. So we have shown how such asymmetries *can* arise, in the way that Stalnaker showed how failures of transitivity in counterfactuals *can* arise. But while we have the distinct facts to classify asymmetrically, we still have the non-logical problem: whence comes the classification? The only suggestion so far is that it comes from Aristotle's concepts of cause and essence; but if so, modern science will not supply it.

4. *The Aristotelian Sieve*

I believe that we should return to Aristotle more thoroughly, and in two ways. To begin, I will state without argument how I understand Aristotle's theory of science. Scientific activity is divided into two parts, *demonstration* and *explanation*, the former treated mainly by the *Posterior Analytics* and the latter mainly by Book II of the *Physics*. Illustrations

²¹ Cf. my "Facts and Tautological Entailments," *The Journal of Philosophy*, vol. 66 (1969), pp. 477-487 and in A. R. Anderson, et al., (ed.), *Entailment* (Princeton, 1975); and "Extension, Intension, and Comprehension" in Milton Munitz (ed.), *Logic and Ontology* (New York, 1973).

in the former are mainly examples of explanations in which the results of demonstration are *applied*; this is why the examples contain premises and conclusions which are not necessary and universal principles, although demonstration is only to and from such principles. Thus the division corresponds to our pure versus applied science. There is no reason to think that principles and demonstrations have such words as "cause" and "essence" in them, although looking at pure science from outside, Aristotle could say that its principles state causes and essences. In applications, the principles may be filtered through a conceptual sieve originating outside science.

The doctrine of the four "causes" (*aitiai*) allows for the systematic ambiguity or context-dependence of why-questions.²² Aristotle's example (Physics II, 3; 195a) is of a lantern. In a modern example, the question why the porch light is on may be answered "because I flipped the switch" or "because we are expecting company," and the context determines which is appropriate. Probabilistic relations cannot distinguish these. Which factors are explanatory is decided not by features of the scientific theory but by concerns brought from outside. This is true even if we ask specifically for an "efficient cause," for how far back in the chain should we look, and which factors are merely auxiliary contributors?

Aristotle would not have agreed that essence is context-dependent. The essence is what the thing is, hence, its sum of classificatory properties. Realism has always asserted that ontological distinctions determine the "natural" classification. But which property is counted as explanatory and which as explained seems to me clearly context dependent. For consider Bromberger's flagpole example: the shadow is so long because the pole has this height, and not conversely. At first sight, no contextual factor could reverse this asymmetry, because the pole's height is a property it has in and by itself, and its shadow is a very accidental feature. The general principle linking the two is that its shadow is a function $f(x, t)$ of its height x and the time t (the latter determining the sun's elevation). But imagine the pole is the pointer on a giant sundial. Then the values of f have desired properties for each time t , and we appeal to these to explain why it is (had to be) such a tall pole.

We may again draw a parallel to counterfactuals.

Professor Geach drew my attention to the following spurious argument: If John asked his father for money, then they would not have quarreled (because John is too proud to ask after a quarrel). Also if John asked and they hadn't quarreled, he would receive. By the usual logic of counterfactuals, it follows that if John asked his father for money, he would receive. But we know that he would not, because they have in fact quarreled. The fallacy is of equivocation, because "what was kept constant" changed in the middle of the monologue. (Or if you like, the aspects by which worlds are graded as more or less similar to this one.) Because science cannot dictate what speakers decide to "keep constant" it contains no counterfactuals. By exact parallel, *science contains no explanations*.

5. *The Logic of Why-Questions*

What remains of the problem of explanation is to study its logic, which is the logic of why-questions. This can be put to some extent, but not totally, in the general form developed by Harrah and Belnap and others.²³

A question admits of three classes of response, *direct answers*, *corrections*, and *comments*. A *presupposition*, it has been held, is any proposition implied by all direct answers, or equivalently, denied by a correction. I believe we must add that the question "Why P , in contrast to X ?" also presupposes that (a) P is a member of X , (b) P is true and the majority of X are not. This opens the door to the possibility that a question may not be uniquely determined by its set of direct answers. The question itself should decompose into factors which determine that set: the *topic* P , the *alternatives* X , and a *request specification* (of which the doctrine of the four "causes" is perhaps the first description).

We have seen that the propositions involved in question and answer must be individuated by something more than the set of possible worlds. I propose that we use the facts that make them true (see footnote 21). The context will determine an asymmetric relation among these facts, of *explanatory relevance*; it will also determine the theory or beliefs which determine which worlds are *possible*, and what is *probable* relative to what.

We must now determine what direct answers are and how they are evaluated. They must be made

²² Cf. Julius Moravcsik, "Aristotle on Adequate Explanations," *Synthese*, vol. 28 (1974), pp. 3-18.

²³ Cf. N. D. Belnap, Jr., "Questions: Their Presuppositions, and How They Can Fail to Arise," *The Logical Way of Doing Things*, ed. by Karel Lambert (New Haven, 1969), pp. 23-39.

true by facts (and only by facts forcing such) which are explanatorily relevant to those which make the topic true. Moreover, these facts must be statistically relevant, telling for the topic in contrast to the alternatives generally; this part I believe to be explicable by probabilities, combining Salmon's and Hannson's account. How strongly the answers count for the topic should be part of their evaluation as better or worse answers.

The main difference from such simple questions as "Which cat is on the mat?" lies in the relation of a why-question to its presuppositions. A why-question may fail to arise because it is ill-posed (P is false, or most of X is true), or because only question-begging answers tell probabilistically for P in contrast to X generally, or because none of the factors that do tell for P are explanatorily relevant in the question-context. Scientific theory enters mainly in the evaluation of possibilities and probabilities, which is only part of the process, and which it has in common with other applications such as prediction and control.

IV. SIMPLE PLEASURES

There are no explanations in science. How did philosophers come to mislocate explanation among semantic rather than pragmatic relations? This was certainly in part because the positivists tended to identify the pragmatic with subjective psychological features. They looked for measures by which to evaluate theories. Truth and empirical adequacy are such, but they are weak, being preserved when a theory is watered down. Some measure of "goodness of fit" was also needed, which did not reduce to a purely internal criterion such as simplicity, but concerned the theory's relation to the world. The studies of explanation have gone some way toward giving us such a measure, but it was a mistake to call this explanatory power. The fact that seemed to confirm this error was that we do not say that we *have* an explanation unless we have a theory which is acceptable, and victorious in its competition with alternatives, whereby we can explain. Theories are applied in explanation, but the peculiar and puzzling features of explanation are supplied by other factors involved. I shall now redescribe several familiar subjects from this point of view.

When a scientist campaigns on behalf of an advocated theory, he will point out how our situation will change if we accept it. Hitherto unsuspected factors become relevant, known relations are revealed to be strands of an intricate web, some terribly puzzling questions are laid to rest as not arising at all. We shall be in a much better position to explain. But equally, we shall be in a much better position to predict and control. The features of the theory that will make this possible are its empirical adequacy and logical strength, not special "explanatory power" and "control power." On the other hand, it is also a mistake to say explanatory power is nothing but those other features, for then we are defeated by asymmetries having no "objective" basis in science.

Why are *new* predictions so much more to the credit of a theory than agreement with the old? Because they tend to bring to light new phenomena which the older theories cannot explain. But of course, in doing so, they throw doubt on the empirical adequacy of the older theory: they show that a pre-condition for explanation is not met. As Boltzmann said of the radiometer, "the theories based on older hydrodynamic experience can never describe" these phenomena.²⁴ The failure in explanation is a by-product.

Scientific inference is inference to the best explanation. That does not rule at all for the supremacy of explanation among the virtues of theories. For we evaluate how good an explanation is given by how good a theory is used to give it, how close it fits to the empirical facts, how internally simple and coherent the explanation. There is a further evaluation in terms of a prior judgment of which kinds of factors are explanatorily relevant. If this further evaluation took precedence, overriding other considerations, explanation would be the peculiar virtue sought above all. But this is not so: instead, science schools our imagination so as to revise just those prior judgments of what satisfies and eliminates wonder.

Explanatory power is something we value and desire. But we are as ready, for the sake of scientific progress, to dismiss questions as not really arising at all. Explanation is indeed a virtue; but still, less a virtue than an anthropocentric pleasure.²⁵

University of Toronto and University of Southern California

Received September 1, 1976

²⁴ Ludwig Boltzmann, *Lectures on Gas Theory*, tr. by S. G. Brush (Berkeley, 1964), p. 25.

²⁵ The author wishes to acknowledge helpful discussions and correspondence with Professors N. Cartwright, B. Hannson, K. Lambert, and W. Salmon, and the financial support of the Canada Council.

VIII. THE PSYCHOLOGY OF JUSTICE IN PLATO

JOHN M. COOPER

ONE important merit of the recent spate of articles addressing the *Republic's* central thesis, that all and only those who are just lead flourishing, fulfilled lives, is to have directed attention anew to Plato's moral psychology—specifically to the links Plato tries to forge between the love for knowledge and understanding and the aversion to (at any rate) much of what would ordinarily be counted as unjust treatment of other people.¹ It has rightly been seen that the core of Plato's response to Thrasymachus' challenge lies in this novel—even paradoxical—thesis about the nature of what a contemporary theorist would call "moral motivation." In this paper I want to examine directly and on its own account Plato's theory of the just person's psychology, leaving aside for the most part the controversies that have loomed so large lately about his justification for using the word "just" (*dikaios*) to describe the person whose psychology this is. The state of mind Plato describes has not, I think, been adequately represented in the recent literature, and my interest is more in what kind of person this is whom Plato calls just, than in whether he is correct in calling him that. My discussion will direct attention more than has been usual in this context to the *Republic's* metaphysics, and less to its political and moral theory, narrowly conceived. I hope thereby to illuminate somewhat the Platonic conception of human perfection and to bring out both the close links between the psychology and the metaphysics of the *Republic* and some of the often unnoticed consequences that Plato's metaphysics has for his psychological theories.

I

Plato develops his account of the psychology of the just person in two principal stages. The later stage, found in *Republic* V–VII, is the more important for present purposes, but before turning to that I must take up briefly the earlier account of book

IV. Plato's argument in the later books cannot be understood properly except in conjunction with the book IV account, aspects of which it is intended to fill out, and some of the recent literature on book IV contains misconceptions about the argument there that must be cleared away before we can proceed.

From the beginning of *Republic* IV it is taken for granted that to be just is to be in a psychological state which shapes and directs one's voluntary actions, one's choices and one's preferences over a very wide range of practical concerns. Plato, for reasons we need not go into now, takes it that such a complex state cannot be adequately understood by treating it as a disposition to exhibit any single, or even any complex set of, independently and non-question-beggingly specifiable sorts of action. Justice must be characterized instead in internal, psychological terms, as a condition of a person's action-, choice- and preference-producing apparatus, specified by reference to interrelationships among the different elements of this apparatus itself. So, having fixed these elements as three, reason, aspiration² and appetite, Plato concludes that justice is the condition of a person in which each of these three plays always and only a certain single role, one for which it is naturally suited. Reason, on its own and without interference from the others, determines how it is best for the person to be and act; aspiration identifies with reason's directives, regarding them as objects to strive for and making the person dissatisfied with himself if through his own fault they are not attained and appetite, restrained and moderated by reason and aspiration, drives him on toward pleasures of eating, drinking, sex and other bodily gratifications when and as, and only when and as, reason approves. When each of these three psychological elements performs just its assigned job, then, Plato holds, the person in that psychological condition is just; and if in any way or to any degree these elements fail to

¹ See especially Richard Kraut, "Reason and Justice in Plato's *Republic*," in E. N. Lee *et al.*, (eds.), *Exegesis and Argument* (Assen, 1973).

² For the idea for this rendering of the Greek *θυμοειδές* I am indebted to J. C. B. Gosling, *Plato* (London, 1973), ch. III.

do their jobs, or attempt to do anything else, such a person is not just. It does not follow, of course, that he is *unjust*. As Socrates admits (472b7-d2), it may well turn out that no one fulfills these conditions perfectly; still, the definition is no less correct and no less useful for that, since we are free, for all practical purposes, to count as just those who sufficiently nearly resemble the person so defined. If it can be shown that only a perfectly just person is truly well-off, and lives a fulfilled life, then the more nearly anyone approximates to this internal condition the closer he will come to leading the perfectly flourishing human life: we need perfect models of just and unjust persons so that "however we find them to be as regards flourishing or its reverse, we will necessarily have to agree also that whoever of ourselves is most like them will have a share [of flourishing] most like theirs," 472c7-d1.³ Strictly, then, only the person who fully attains precisely this psychological condition is just.

Even so, it is obvious enough that Plato has presented no sufficiently determinate theory of justice until he explains more fully what the role of reason is to be and how it is to go about performing it. He explains at some length in books III and IV what functions the other two parts are to perform, but with the simple idea, which is all I have enunciated so far, that reason is to rule—to determine what is to be done and to put these determinations into effect—one has at best a formal criterion from which no substantive choices and preferences can be derived. Plato is deliberately vague in book IV on this point, since he is not yet in a position to define the role of reason more precisely; but he is not nearly so vague even in IV as he is sometimes thought to be. For in the full statement of what the function of reason is, Plato says not merely that reason is to rule but that it is to rule *with wisdom*: "Does not it belong to the rational part to rule, being wise and exercising forethought on behalf of the entire soul, and to the principle of high spirit to be subject to this and its ally?" (441e4-6, tr.

Shorey). And a little farther on he characterizes wisdom: "But [we call an individual] wise by that small part that ruled in him and handed down these commands, by its possession in turn within it of the knowledge (*ἐπιστήμη*) of what is beneficial for each and for the whole, the community composed of the three" (442c5-8, tr. Shorey). So a person's reason is not doing its job unless three conditions are met: first, it *knows* what it is best for him to do; second, it makes all its decisions in the light of this knowledge; and third, its decisions are effective.⁴ The main contribution of books V-VII is indeed to work out a theory of what this knowledge consists of; but already in IV Plato is explicit that unless one's reason has this knowledge, whatever exactly it amounts to, it is not performing its work fully and adequately.

It follows from what has been said so far that on Plato's account no one is just, strictly speaking, who does not have *knowledge* of what is best to do. True belief is not sufficient for an individual's justice in the *Republic*, nor, I think, though this is a more complicated question, for any other virtue of individuals.⁵ It is true that Plato conspicuously defines a *city's* courage (429a-430c) in terms of the deep-dyed belief (not knowledge) of the soldier-class in the correctness of the laws and institutions of the city they serve. But it is noteworthy that he denies that this condition of belief makes the soldiers themselves brave, except in a qualified sense: it makes their city brave, but it gives them, as Socrates cautions, only "civic bravery" (*πολιτικὴ ἀνδρεία*, 430c3), not bravery *tout court*. It makes them consistently do the things one expects of brave citizens—they are fearless and selfless soldiers and police officers—but they are far from having the philosopher's pervasive strength of character, that will not let one rest content until one has achieved the fullest possible understanding of what is good for oneself and why it is so.⁶ Similarly the city's justice is defined as the condition in which each class sticks to its own social work (432d-435a), but

³ On the same principle Plato justifies the moral and political subjection of everyone else in his ideal state to the philosophers: by being so subject they approximate as closely as is possible for them to the state of soul possessed by the philosophers, and so also to the condition of flourishing enjoyed by them. See 590c-d.

⁴ Thus it is clear that Plato does not work even in *Rep.* IV with what Kraut calls the "non-normative" conception of reason's rule (Kraut, *op. cit.*, pp. 208-211).

⁵ This has been denied by Gregory Vlastos, "The Argument in the *Republic* that 'Justice Pays'," *The Journal of Philosophy*, vol. 65 [1968], pp. 665-674, and its revision and expansion, "Justice and Happiness in the *Republic*" in Gregory Vlastos (ed.), *Plato II* (Garden City, 1971), pp. 66-95; see esp. the latter, pp. 92-94.

⁶ See 503b7-504d3: to give the best possible account of the virtues one must take a longer route than was taken in describing them in book IV, one which makes explicit reference to knowledge of the good. Notice especially 503c: we require as rulers those who do not shy away like cowards (*ἀποδευλιάσει*) in the face of difficult intellectual tasks (and, to the same effect, 535b5-g); see also the martial and gymnastic metaphors of 534b8-c3.

this does not mean, nor does Plato anywhere suggest, that people who do stick to their social work thereby show themselves to *be* just. They do what is just, but whether they *are* just is obviously not settled by pointing to their behavior alone. "Doing one's own social work" is presented only as a description of just action. It is not presented as part of the definition of the condition of justice itself—except, of course, the justice of a city, which is a different matter entirely.⁷ Here Plato does not even allow an analogue of "civic bravery," and for a very good reason: in defining the city's justice he has said nothing about the internal condition(s) of the citizens, which cause(s) them to act justly, so he has not provided any ground here, as he has in the case of bravery, for attributing even a reduced form of the virtue to the individuals themselves. Once these points are duly noted, one sees that Plato consistently restricts justice, as a virtue of individuals, to those who possess within themselves *knowledge* of what it is best to do and be.⁸

II

Clearly, then, one cannot know what the just person according to Plato is like until one has some grasp, at least, of the nature and substance of this knowledge. For a person's considered conception of how it is best to be and act inevitably colors his attitudes to himself and to other people, as well as establishing or altering more particular attach-

ments and antipathies — and especially so in the case of someone who is so structured that there is no possibility of opposition or resistance to the effective and total adoption of the recommendations of his reason. And if one wants to know what the state of mind of Plato's just person would be, one will have to know something about how he thinks and feels in these respects.

In book IV itself, as already remarked, little is said about what the just person knows. It is noteworthy, however, that the wisdom he is said to have is described merely as knowledge of what is advantageous to himself (442c5–8). Yet earlier in the book the same name, *σοφία*, is given to the knowledge, possessed by the rulers of the ideal city, "which takes counsel about the city as a whole as to how it would best order its relations to itself and to other cities" (428c12–d3; cf. 429a1–3). Unless Plato is guilty here of a gross equivocation, he must suppose that it is one and the same knowledge that knows both these things.⁹ No doubt part of his reason for supposing this derives from his conviction that it is only in cities that human beings realize their full nature; but there is a more general reason as well. For both of these accomplishments essentially involve knowledge of what is good (for the city, for the individual); and you cannot, Plato thinks, know any such particular good without knowing the good-itself: "Whoever cannot mark off in his discourse the form of the good, separating it from all other things . . . you will say knows

⁷ Vlastos concedes this point in the later version of his paper ("Justice and Happiness in the *Republic*," *op. cit.*, p. 79), but insists that according to Plato a city can be just only if its citizens are just. A moment's reflection shows this to be an extraordinary view for anyone to hold, and especially so for Plato who, as we have seen, very clearly requires wisdom and knowledge of anyone who is just: surely he cannot have slipped even temporarily into allowing *all* the citizens of his city to have this knowledge? In fact the passage Vlastos cites for this view (435e1–436a3) says nothing of the kind. It does not say that in general "moral attributes," as Vlastos has it, come to be predicated of a city only by the antecedent possession of the same attributes by its inhabitants. Plato's point is a more limited one, in accordance with his aims in the passage in which it occurs. Socrates is suggesting, as he will presently argue, that each of us contains three psychological elements corresponding to the three classes of the state. An indication of this, he thinks, is the fact that whole nations are commonly thought to have predominant characters (*ἡθῆ*, e2)—the Scythians, for example, to be high-spirited, the Athenians devoted to learning and the Phoenicians fond of money. These characters must attach to these nations by transference from their members (and given that the three characters in question seem to suggest the prominence of the ends of one or another of three distinct parts of the soul, this usage therefore implies that human beings do have the three distinct parts in question). What is true of such predominant national characters need not be true of other properties, not even all "moral attributes"; and Plato nowhere says it is (544d–545b, the only other passage where any similar point is made, applies it to the same restricted set of attributes as we find in this passage). So far as the moral virtues are concerned, as I argue in the text, Plato says different things about different ones: wisdom belongs to the city because the rulers are wise (cf. the repeated emphasis on *ἐπιστήμη*, 428b6, c11, d8, e8, 429a1–3), bravery because the soldiers have a qualified form, at least, of bravery; justice and moderation, however, though they belong to the city in each case because of *some* property of the citizens, do not devolve upon it from the citizens' being even in a qualified way just and moderate. This seems to me a perfectly sensible view; it is, in any event, for better or worse, what Plato thought.

⁸ Hence for Plato only accomplished philosophers can be just. Things would have been different, and better, if like Aristotle he had conceived of a kind of practical knowledge substantially independent of the use of theoretical and scientific powers. He was precluded from doing this by the status of the good-itself in his theory; see below, sect. III.

⁹ Aristotle, *NE* V, 1141b24–25, *ἔστι δὲ καὶ ἡ πολιτικὴ καὶ ἡ φρόνησις ἡ αὐτὴ μὲν ἕξις, τὸ μὲντοι εἶναι οὐ τῶν αὐτῶν, κτλ.*, simply reformulates in his own terminology this view of Plato's about these forms of *σοφία*.

neither the good-itself nor any other good" (534b8-c5). To be sure, merely knowing the good will not by itself suffice—at any rate not for knowing how best to run a city. This is why Plato insists that before coming to know the good his rulers must spend fifteen years getting experience in city management (539e2-5). Still, the only part of what either sort of person knows that counts, on a strict view, as knowledge (the sort that encompasses understanding, of why things are the way they are and how they ought to be) is knowledge of the good-itself. Without this, all the experience in the world would be useless. So what Plato describes in book IV now as knowledge of how one ought to live oneself, now as knowledge of how a city should conduct its affairs, is in the end nothing less than knowledge of the μέγιστον μάθημα, the good-itself.

Thus we are led directly up to the central metaphysical theories of the *Republic*. This part of the dialogue is extremely difficult to interpret, since Socrates himself refuses to speculate about what the good-itself might be (506b2-e5) and for the most part offers only the treacherous analogies and parables of sun, line and cave to convey those aspects of it that he is willing to hazard an opinion on. But without claiming to offer any proof that this is what Plato means, and avoiding scholarly controversy so far as possible, I risk the following summary of what is relevant in Plato's views on the good-itself for our present inquiry.

First, pending a qualification to be entered directly, the good-itself is a good thing, over and above the good things of this world:¹⁰ over and above, that is, both individual good things like particular persons and events, and such things as a quiet, studious life or a cool drink on a hot day which one might speak of as good things, though they are not individuals but classes of individuals.¹¹ Unlike these other good things, however, it is not merely a good something-or-other, or good as such and such, or good for so and so or from such and such a point of view. To use Aristotelian language, one could say its essence is to be good; it is not, like

every other good thing, essentially something else (a meal, a person) that, for one reason or another, happens to be good (is accidentally good).¹² Its goodness is not, therefore, diluted and compromised by being mingled with and made dependent on other features of things, as is true of every other good. Thus its goodness is pure, as that of no other good is. Furthermore, it is the only perfect good. Every other good, being good only in some respect or relation or from some point of view, is also not so good, or even quite bad, in some other ways.¹³ No other good can possibly be as good as this one: the very best imaginable human life, for example, will still not be as good as this is, and so not as good as it is possible to conceive something as being.

Given the purity of the good-itself, and the impurity of every other good, it should be apparent why Plato says that one cannot know any other good except by knowing this one. Knowledge here, as usual in Plato, is taken to entail understanding, so that to know, e.g., a good life is not just to know which life is good, and from what points of view or in what circumstances, and so on, but precisely to understand what this goodness itself is that one attributes to it with these qualifications. To understand that, however, is to conceive a good that stands apart from all such qualifications—in short, to form one's conception of goodness on the good-itself.

Given the perfection of the good-itself, and the imperfection of every other good, it is apparent why Plato insists that adequate practical thinking requires constantly keeping this good uppermost in one's mind. Everyone wants what is best, but the only sure criterion of goodness in any other thing is the degree to which it approximates the goodness of this perfect good. Any other criterion of goodness, based, say, merely on some conception of human needs and interests, cannot fail to be inadequate, since it must represent as the ideal, as the final achievement, something that cannot be that. Clarity and truth require that all partial standards be seen and treated as such.

¹⁰ I do not follow Vlastos in interpreting Plato's claims that the *F*-itself is *F* as in most cases predicating *F*-ness only of the *F* things of this world—at any rate this move cannot be made with the good-itself. See his "The Unity of the Virtues in the *Protagoras*" and "An Ambiguity in the *Sophist*" in his *Platonic Studies* (Princeton, 1973).

¹¹ I owe a debt here to N. R. Murphy, *The Interpretation of Plato's Republic* (Oxford, 1951), ch. VI, and to J. C. B. Gosling, "Republic Book V: τὰ πολλὰ καλὰ, etc.," *Phronesis*, vol. 5 (1960), 116 ff., though my interpretation is not identical, I think, with any of the various ones they seem to have in mind.

¹² Hence it is misleading to say, as I just did, that it is a good thing. "Good" belongs to it in much the way that "man" does to me, and it would be misleading or nonsensical to call me a man thing: that would invite the inappropriate question, "Man what?" So the good is a good, but, Plato will insist (to Aristotle's exasperation, cf. *Post. An.* 83a30-35) not a good thing.

¹³ For a defense of this interpretation of the perfection of Platonic forms, see Alexander Nehamas, "Plato on the Imperfection of the Sensible World," *American Philosophical Quarterly*, vol. 12 (1975), pp. 105-117.

So much for what one might call the functional properties of the good-itself. How about its substance or nature? Here Socrates is deliberately least informative. One may, however, render this curious entity more concrete by thinking of it somehow or other as a perfect example of rational order, conceived in explicitly mathematical terms: a complex, ordered whole, whose orderliness is due to the mathematical relationships holding among its parts. This is, I think, implied by the mathematical nature of the higher education that Plato prescribes in *Republic* VII (524d ff.) as preparing the way for dialectic, which itself culminates in the knowledge of this entity, and by the role which he assigns to knowledge of the good as completing and supporting the mathematical sciences of arithmetic and geometry (511c6-d1, 533b6-d1).¹⁴ It should be recalled that the last stage of the preliminary studies, mathematical harmonics, already encompasses the idea that certain numerical relationships are in themselves harmonious and itself provides some explanation for why they harmonize (531c3-4); notice, too, that Socrates recommends the study of music for this reason especially, saying that this makes it "useful for the investigation of the beautiful and the good" (c6-7).¹⁵ Thus mathematics and philosophy merge at their borders, and mathematics itself is capable, on Plato's account, of handling and partially explicating evaluative notions.

III

If this account of the good-itself is right, what must be the state of mind of the person who has a firm grasp on it? First of all, as Plato emphasizes in many places, he must be a lover of learning (a *φιλομαθής*). This is so, to begin with, because only someone who felt a natural affinity for reading and thinking and abstract discussion and had learned to value these things more highly than almost anything else could ever have persevered through the rigorous course of training that must be undergone to achieve it, with the attendant restriction of bodily gratification and curtailment of other sorts of pleasant pursuit. This knowledge is obviously not easily won, and if justice requires having it, and only a studious sort of person, who cares relatively little for other pursuits in comparison with intellectual ones, can achieve it, then no one can be just who is not that kind of studious person.

This argument depends, of course, on somewhat doubtful causal connections, and it is worth noting that much the same conclusion follows more directly and firmly from the very nature of what the just person according to Plato knows. He knows the good-itself and therefore whatever he values he values strictly in the light of a comparison between that thing and the good. Whatever exhibits more fully and perfectly the sort of rational order that the form of the good possesses as its essence he values more highly than other things. Naturally, it would be rash to claim to know how he would rank things in the light of this knowledge. But it does seem inevitable that he would find in intellectual work generally, and contemplation of the good-itself in particular, the most nearly adequate instances of rational order in the natural world. This is so because pure, i.e., theoretical, rational thought is to the greatest degree possible completely determined by the requirements of rational order itself: by contrast, desires, material objects and everything else that is a mixture or combination of thought and bodily things and events, even when they do exhibit rational order, must remain a joint product of reason and something else. What such good things are like is largely determined, not by the demand for orderliness, but by the nature of these things themselves. The best pure thinking, on the other hand, is almost wholly the product of rational orderliness itself. Thus the contemplation of the good, and in lesser degree all other abstract scientific thought, since in these activities the impulse for rational order confronts less in the way of alien material to work upon, must be the most perfect earthly embodiments of good; and just men and women would know this and make this knowledge effective in their own lives. Hence they would prefer this kind of thinking to everything else; in it they would be closest to the good-itself, the only perfectly good thing there is.

On the other hand, it does not follow that they would always choose to live a contemplative's life, retired among their mathematical books and constantly engaging in philosophical discussions. Perhaps one would choose this if all he cared about was *his* realizing the good so far as possible. But a just person is a devotee of *the* good, not *his own* good; and these are very different things. Knowing the good, what he wants is to advance the reign of rational order in the world as a whole, so far as by

¹⁴ It also accommodates those later reports of a lecture *On the Good* that, to the surprise of some of its hearers, had to do with mathematics and little else (cf. Aristoxenus, *Harm.* 2.20.16-31.3).

¹⁵ On this see Gosling, *Plato, op. cit.*, ch. VII.

his own efforts, alone or together with others, he can do this. He recognizes a single criterion of choice: What, given the circumstances, will be most likely to maximize the total amount of rational order in the world as a whole? And here he has a wide arena for possible activity. He can not only impose rational order on his own soul, thinking rational thoughts and satisfying rationally controlled appetites of his own; he can help to bring rational order to the souls of other individuals, and to their social life. To be sure, under certain conditions (for example, perhaps, those characteristic of Plato's own time: cf. 496c5-e2) he might find it rational not to work for the improvement of anything except himself. If conditions make it impossible for his efforts to bear fruit, then his time would be wasted in interesting himself in others: the reign of rational order would not thus be advanced, or not significantly enough to compensate for the loss of pure thinking the world would have to suffer in consequence of his taking time off from his studies. On the other hand, conditions need not always be so bad as that. There may be others intelligent enough to reach the higher intellectual realms themselves and willing to learn from him how to achieve them; and society itself may be ready, if he and others who know the truth work together, to accept their recommendations for its improvement. Under such conditions it would seem that the interests of rational order would be better served by his devoting himself, some of the time, to private teaching and to co-operation with other intellectuals in bettering the condition of mankind generally—in addition, of course, to doing a substantial amount of philosophical work of his own. This combination of activities, it seems plausible to say, would under such conditions constitute his best answer to the question: How can I maximize the total amount of rational order in the world as a whole?

Hence I believe Plato is perfectly entitled to the answer his Socrates gives to Glaucon's complaint (519d8-9) that it is unjust to make those who in the ideal city have reached the goal of intellectual culture leave off intellectual work and spend some of their time putting their knowledge to practical use in the improvement of the lives of others. Socrates answers, in part, that, on the contrary, it would not be fair of them to refuse, seeing that they have themselves exploited the city's institutions for their own intellectual benefit (520a6-c1): and, Glaucon now agrees, just people, such as these men and women are, will not refuse to do what is fair (*δίκαια*

γὰρ δὴ δίκαιους ἐπιτάξομεν, e1). In saying this Socrates is transparently appealing to common sense principles of justice, which he elsewhere (see 538d6-539a1, esp. the implications of e6, *τὰ ἀληθῆ μὴ εὐρίσκειν*) rejects as not true. But there is no cause for alarm, since the purer principles Plato himself has been espousing lead to the same conclusion. That these philosophers are just means, on Plato's account, that they know the good-itself and act always with a view to advancing rational order in the whole world. And given the conditions prevailing in the ideal state it does seem very reasonable to think that they would, on their own principles, opt for the mixed political and intellectual life which Socrates insists on.

On the other hand, it is right to emphasize the other part of Socrates' reply to Glaucon as well. For, he insists, they will go down into the cave unwillingly (*ὡς ἐπ' ἀναγκαῖον*, as Glaucon puts it, 520e1; cf. 540b4-5), that is to say not thinking that (of all things!) the good is to be found in the activities of ruling (cf. 521a5-6, b4-5) and recognizing that there is a life better than the political one (520e4-5, 521b9). They deliberately and freely (520d6-7) choose a life for themselves that is less good than a more singlemindedly intellectual life, of which however they are individually capable. The life they renounce is better because, taken just by itself, the activities it contains exhibit, on balance, a greater amount of rational order than do the combination of activities in the alternative. Furthermore, this life can be said to be in general the ideal best for any human being. There is no activity that comes closer to perfect embodiment of the good than contemplation, with understanding, of the good-itself and the other forms; and given, as Plato makes very clear, that human beings can come into existence who can engage in this activity more or less uninterruptedly, it follows that no other human life could be so good as this one. Hence if the degree of the philosopher's *εὐδαιμονία* is judged by comparison with this ideal, Plato's philosophers will settle for a less flourishing existence than they might have had (519e1-520a4; cf. 420b4-8, 465e4-466a6). On the other hand, as previously noted, a true philosopher never concerns himself merely with his own good. His ultimate end is to improve not just the small part of the world that is constituted by his own life, but the whole of it, this part taken together with all the rest. So if a philosopher opted, under the conditions Plato envisages, to ignore everyone else in order to make his own life realize the good more perfectly and

fully, he would fail to achieve the goal he was aiming at as nearly as he might have done. And if the degree of one's *eudaimonia* is measured by how close one comes to realizing one's ultimate end, such a philosopher would be less *eudaimon* than he would have been by living the mixed political and intellectual life Socrates and his interlocutors were urging. And if one further supposes, as Plato certainly does, that human nature is such that left to themselves most people would always lead *very* disordered lives, it seems fair to say that any philosopher who ever opts for the mixed life will actually be more *eudaimon* than any who opts for the purely intellectual life: any philosopher would always *prefer* the mixed life, and he would recognize any situation in which it was rational for him to choose a life of pure contemplation instead as one where external conditions alone, by preventing his efforts on others' behalf from bearing fruit, forced him to settle for less than he had wished to achieve on behalf of the good-itself. So, as Socrates intimates (420b4-5), his philosophers, making the choice to spend some of the time in the cave, but most of it in the world above, would be the happiest and most flourishing men there ever in fact can be.

This shows beyond any reasonable doubt that Plato's just man is no egoist, in any acceptable sense of this term. Not only does he not do everything he does out of concern for his own good, he never does anything for this reason. Even where he acts to benefit himself, recognizing that he does so, his reason for acting is that the good-itself demands it. That *his* good demands it is strictly irrelevant. By the same token, at no time does he act to benefit others out of regard for them and concern for *their* good, just because it is theirs. Again, he confers all benefits out of regard for the good-itself, not out of regard for these more immediate human beneficiaries. As Plato makes clear, the philosopher turns to this work with regret, *ὡς ἐπ' ἀναγκάϊον*, and not with any independent attachment to these activities themselves or to those at whose good they are directed. For him there is no pleasure or interest in them for their own sakes. This is, I think, the direct consequence of the role played in Plato's theory by the form of the good. It is the ultimate object of pursuit, yet lies outside the world. Hence no worldly thing or activity can, because of its own properties, because of what *it* is, interest the just

man; anything interests him only as a means of coming nearer to the good-itself. Certainly this must be true for most ordinary human activities and interests, involving as they do the expression of appetites and emotions. To this rule there is a single exception, the activity of rational contemplation itself. Since that is the summit of human experience anyone who takes pleasure in it and loves it for its own sake will not be in danger of compromising his pursuit of the ultimate end itself. Because the just person's decision whether to use, or continue to use, his knowledge for the benefit of others is contingent on factors outside his control, he must avoid developing sentiments of attachment to other people and to the normal round of social and political life. But there is no possibility that someone who can contemplate the forms with understanding could ever be forced by circumstances to give up this activity entirely. This worldly attachment, therefore, is firm in a way that no other can be. This means that though the just man in Plato is no egoist, and no altruist either, but a sort of high-minded fanatic; still, his fanaticism is such as to allow him permanent and deep interests in one part of his own good, whereas he cannot similarly have any permanent and deep attachment to any other person's good. This is the only concession that one must grant to the common charge that Plato's just man is a covert, or not so covert, egoist.

I said a moment ago that all this is the consequence of the role played in Plato's theory by the form of the good. Aristotle was right, I think, to insist most emphatically that there is no such thing as a good-itself and that even if there were one ought not to direct one's practical thinking at it as ultimate end. At any rate, the consequence of this insistence is something his own moral sensitivity seems to have made him very reluctant to give up: by placing the object of ultimate pursuit in the world Aristotle is able to make room for a wide variety of ordinary human pursuits, interests and attachments as themselves permanent constituents of the final good.¹⁶ He could therefore present a moral psychology that corresponds much more closely than Plato's does to ordinary moral experience. No doubt Plato would not regard this as a defect in his own theory or a point in favor of Aristotle's; but the fact itself is worth reflecting on.

University of Pittsburgh

Received September 1, 1976

¹⁶ Cf. my *Reason and Human Good in Aristotle* (Cambridge, Mass., 1975), esp. ch. II, and "Friendship and the Good in Aristotle," forthcoming in *The Philosophical Review*.

IX. THE BASIC STRUCTURE AS SUBJECT

JOHN RAWLS

I

AN important assumption of my book *A Theory of Justice*¹ is that the basic structure of society is the primary subject of justice. By the basic structure is meant the way in which the major social institutions fit together into one system, and how they assign fundamental rights and duties and shape the division of advantages that arises through social cooperation. Thus the political constitution, the legally recognized forms of property, and the organization of the economy, all belong to the basic structure. I held that the first test of a conception of justice is whether its principles provide reasonable guidelines for the classical questions of social justice in this case.

In my book I did not consider in any detail why the basic structure is to be taken as the primary subject. I left this to be gathered from various remarks made while discussing other matters. Here I shall try to remedy this lack. Of course, it is perfectly legitimate at first to restrict inquiry to the basic structure. We must begin somewhere, and this starting point may turn out to be justified by how everything works out. But certainly we would like to find a more illuminating answer than this; and moreover one that draws upon the special features of the basic structure in contrast with other social arrangements, and connects these features with the particular role and content of the principles of justice themselves. I aim to present an explanation that meets these conditions.

Now a social contract is an agreement (1) between all rather than some members of society, and it is (2) between them as members of society (as citizens) and not as individuals who hold some particular position or role within it. In the Kantian form of this doctrine, of which the conception of justice as fairness is an example, (3) the parties are regarded as, and also regard themselves as, free and equal moral persons; and (4) the content of the agreement is the first principles that are to regulate the basic structure. We take as given a short list of conceptions of justice developed by the tradition of

moral philosophy and then ask which of these the parties would acknowledge, when the alternatives are thus restricted. Assuming that we have a clear enough idea of the circumstances necessary to insure that any agreement reached is fair, the content of justice for the basic structure can be ascertained, or at least approximated, by the principles that would be agreed to. (Of course, this presupposes the reasonableness of the tradition; but where else can we start?) Thus pure procedural justice is invoked at the highest level: the fairness of the circumstances transfers to fairness of the principles adopted.

I wish to suggest the following: first that once we think of the parties to a social contract as free and equal (and rational) persons, then it is natural to take the basic structure as the primary subject. Second, that in view of the distinctive features of this structure, the initial agreement, and the conditions under which it is made, must be understood in a special way that distinguishes this agreement from all others; and doing this allows a Kantian view to take account of the fully social nature of human relationships. And finally, that while a large element of pure procedural justice transfers to the principles of justice, these principles must embody an ideal form for the basic structure in the light of which ongoing institutional processes are to be constrained and the accumulated results of individual transactions continually corrected.

II

Several lines of reasoning point to the basic structure as the primary subject of justice. One is the following: suppose we begin with the initially attractive idea that society should develop over time in accordance with free agreements fairly arrived at and fully honored. Straightway we need an account of when agreements are free and the social circumstances under which they are reached are fair. In addition, while these conditions may be fair at an earlier time, the accumulated results of

¹ Cambridge, Mass., 1971.

many separate ostensibly fair agreements, together with social and historical contingencies, are likely as time passes to alter institutions and opportunities so that the conditions for free and fair agreements no longer hold. The role of the basic structure is to secure just background conditions against which the actions of individuals and associations take place. Unless this structure is appropriately regulated and corrected, the social process will cease to be just, however free and fair particular transactions may look when viewed by themselves.

We recognize this fact when we say, for example, that the distribution resulting from voluntary market transactions (even should all the ideal conditions for competitive efficiency obtain) is not, in general, fair unless the antecedent distribution of income and wealth as well as the structure of the system of markets is fair. The existing wealth must have been properly acquired and all must have had fair opportunities to earn income, to learn wanted skills, and so on. Again, the conditions necessary for background justice can be undermined, even though nobody acts unfairly or is aware of how the conjunction of contingencies affects the opportunities of others. There are no feasible rules that it is practicable to impose on economic agents that can prevent these undesirable consequences. These consequences are often so far in the future, or so indirect, that the attempt to forestall them by restrictive rules that apply to individuals would be an excessive if not impossible burden. Thus we start with the basic structure and try to see how this system itself should make the corrections necessary to preserve background justice.

III

A second reflection points in the same direction. Consider the situation of individuals engaged in market transactions. We have seen that certain background conditions are necessary for these transactions to be fair. But what about the nature of individuals themselves: how did they get to be what they are? A theory of justice cannot take their final aims and interests, their attitude to themselves and their life, as given. Everyone recognizes that the form of society affects its members and determines in large part the kind of persons they want to be as well as the kind of persons they are. It also limits people's ambitions and hopes in different ways, for they will with reason view themselves in part according to their place in it and take account of the means and opportunities they can realistically

expect. Thus an economic regime is not only an institutional scheme for satisfying existing desires and aspirations but a way of fashioning desires and aspirations in the future.

Nor, similarly, can we view the abilities and talents of individuals as fixed natural gifts, even if there is an important genetic component. These abilities and talents cannot come to fruition apart from social conditions and as realized they always take but one of many possible forms. An ability is not, for example, a computer in the head with a definite measurable capacity unaffected by social circumstances. Among the elements affecting the realization of natural capacities are social attitudes of encouragement and support and the institutions concerned with their training and use. Thus even a potential ability at any given time is not something unaffected by existing social forms and particular contingencies over the course of life up to that moment. So not only our final ends and hopes for ourselves but our realized abilities and talents reflect, to a large degree, our personal history, opportunities, and social position. What we might have been had these things been different, we cannot know.

Finally, both of the preceding considerations are strengthened by the fact that the basic structure most likely contains significant social and economic inequalities. These I assume to be necessary, or else highly advantageous, in maintaining effective social cooperation; presumably there are various reasons for this, among which the need for incentives is but one. Even if these inequalities are not very great, they seem bound to have a considerable effect and so to favor some over others depending upon their social origins, their realized natural endowments, and the chance coincidences and opportunities that have come their way. The basic structure includes inequalities between certain starting-places, so to speak, and this feature, together with the earlier observations, prompts us to take this structure as the primary subject.

IV

In the conception of justice as fairness the institutions of the basic structure are viewed as just provided they (reasonably) satisfy the principles that free and equal moral persons, in a situation that is fair between them, would adopt for the purpose of regulating that structure. The main two principles read as follows: (1) Each person has an equal right to the most extensive scheme of equal

basic liberties compatible with a similar scheme of liberties for all. (2) Social and economic inequalities are permissible provided that (a) they are to the greatest expected benefit of the least advantaged; and (b) attached to positions and offices open to all under conditions of fair equality of opportunity.

Let us consider how the special features of the basic structure affect the conditions of the initial agreement and hence the content of these principles. Now by assumption the basic structure is the all-inclusive social system that determines background justice; so any fair situation between individuals conceived as free and equal moral persons must be one that suitably evens out the contingencies within this system. Agreements reached when people know their present place in an ongoing society would be influenced by disparate social and natural contingencies. The principles adopted would then be selected by the historical course of events that took place within that structure. We would not have gotten beyond social happenstance in order to find an independent standard.

It is also clear why, when we interpret the parties as free and equal moral persons, they are to reason as if they know very little about themselves (referring here to the restrictions of the veil of ignorance). For to proceed otherwise is still to allow the disparate and deep contingent effects of the social system to influence the principles adopted; and this is true even if the parties have no particular information about themselves but only general facts about their own society (which is perhaps all that a condition of impartiality requires). When we as contemporaries are influenced by a general description of the present state of society in agreeing how we are to treat each other, and those generations that come after us, we have not yet left out of account the accidents of the basic structure. And so one arrives at the thicker rather than the thinner veil of ignorance: the parties are to be understood so far as possible solely as moral persons, that is, in abstraction from all those contingencies that the basic structure over time has shaped and influenced; and to be fair between them, the initial situation must situate them equally for as moral persons they are equal: the same essential properties qualify each.

Finally, the social contract must be regarded as hypothetical. Of course, any actual agreement is liable to the distortions just noted; but in any case, historically valid compacts, were such to exist, would have but limited force and could not serve as the basis of a general theory. Equally decisive is

the fact that society is a system of cooperation that extends over time: it is cooperation between generations and not just cooperation among contemporaries. If we are to account for the duties and obligations between generations, there is no clear way to do this in a contract view without interpreting the initial agreement as hypothetical. The correct principles for the basic structure are those that the members of any generation (and hence all generations) would agree to as the ones their generation is to follow and as the principles they would want other generations to have followed and to follow subsequently, no matter how far back or forward in time.

Once we note the distinctive role of the basic structure and abstract from the various contingencies within it to find an appropriate conception of justice to regulate it, something like the notion of the original position seems inevitable. It is a natural extension of the idea of the social contract when the basic structure is taken as the primary subject of justice.

V

The essential point is the distinctive role of the basic structure: we must distinguish between particular agreements made and associations formed within this structure, and the initial agreement and membership in society as a citizen. Consider first particular agreements: typically these are based on the parties' known (or probable) assets and abilities, opportunities and interests, as these have been realized within background institutions. We may assume that each party, whether an individual or an association, has various alternatives open to them, that they can compare the likely advantages and disadvantages of these alternatives, and act accordingly. Under certain conditions someone's contribution to a joint venture, or to an on-going association, can be estimated: one simply notes how the venture or association would fare without that person's joining, and the difference measures their worth to the venture or association. The attractiveness of joining to the individual is similarly given by a comparison with their opportunities. Thus particular agreements are reached within the context of existing and foreseeable configurations of relationships as these have been and most likely will be realized within the basic structure; and it is these configurations that give meaning to contractual calculations.

The context of a social contract is strikingly different, and must allow for three facts, among others: namely, that membership in our society is given, that we cannot know what we would have been like had we not belonged to it (perhaps the thought itself lacks a sense), and that society as a whole has no ends or ordering of ends in the sense that associations and individuals do. The bearing of these facts is clear once we try to view the social contract as an ordinary agreement and ask how deliberations leading up to it would proceed. Since membership in their society is given, there is no question of the parties comparing the attractions of other societies. Moreover, there is no way to identify potential contribution to society as an individual not yet a member of it; for this potentiality cannot be known and is, in any case, irrelevant to their present situation. Not only this, but from the standpoint of society as a whole *vis-à-vis* any one member, there is no set of agreed ends by reference to which the potential social contributions of an individual could be assessed. Associations and individuals have such ends, but not a well-ordered society; although it has the aim of giving justice to all its citizens, this is not an aim that ranks their expected contributions and on that basis determines their social role. The notion of an individual's contribution to society as itself an association falls away. It is necessary, therefore, to construe the social contract in a special way that distinguishes it from other agreements.

In the conception of justice as fairness this is done by constructing the notion of the original position. This construction must reflect the fundamental contrasts just noted and it must supply the missing elements so that an appropriate agreement may be reached. Consider in turn the points in the preceding paragraph. First, the parties in the original position suppose that their membership in the society is given. This presumption reflects the fact that we are born into our society and within its framework realize but one of many possible forms of our person; the question of entering another society does not arise. The task is to agree on principles for the basic structure of the society of one's birth. Second, the veil of ignorance not only establishes fairness between equal moral persons, but by excluding information about the parties' actual interests and abilities, it represents the fact that apart from our place and history in a society,

even our potential abilities cannot be known and our interests and character are still to be formed. Thus, the initial situation suitably recognizes that our nature apart from society is but a potential for a whole range of possibilities. Third and finally, there is no social end except that established by the principles of justice themselves, or else authorized by them; but these principles have yet to be adopted.

Nevertheless, although the calculations that typically influence agreements within society have no place, other aspects of the original position provide the setting for rational deliberation. Thus the alternatives are not opportunities to join other societies, but instead a list of conceptions of justice to regulate the basic structure of one's own society. The parties' interests and preferences are given by their desire for primary goods. Their particular final ends and aims indeed are already formed, although not known to them; and it is these already formed interests that they seek to protect by ranking conceptions on the basis of their preference (in the original position) for primary goods. Finally, the availability of general social theory gives a sufficient basis for estimating the feasibility and consequences of the various conceptions of justice. These aspects of the original position allow us to carry through the idea of the social contract despite the unusual nature of this agreement.

VI

I now point out three ways in which the social aspect of human relationships is reflected in the content of the principles of justice themselves. First, the difference principle (which governs economic and social inequalities) does not distinguish between what is acquired by individuals as members of society and what would have been acquired by them had they not been members. Indeed, no sense can be made of the notion of that part of an individual's social benefits that exceed what would have been their situation in another society or in a state of nature. We can, if we like, in setting up the argument from the original position, introduce the state of nature in relation to the so-called no-agreement point.² This point can be defined as general egoism and its consequences, and this can serve as the state of nature.

² See *A Theory of Justice* (*op. cit.*), pp. 136, 147; see also p. 80.

But these conditions do not identify a definite state. All that is known in the original position is that each of the conceptions of justice available to the parties has consequences superior to general egoism. There is no question of determining anyone's contribution to society, or how much better off each is than they would have been had they not belonged to society and then adjusting the social benefits of citizens by reference to these estimates. Although we may draw this kind of distinction for agreements made within society, the requisite calculations for principles holding for the basic structure itself have no foundation. Neither our situation in other societies, nor in a state of nature, has any role in comparing conceptions of justice. And clearly these notions have no relevance at all in the application of the principles of justice.

Second, and related to the preceding, the two principles of justice regulate how entitlements are acquired in return for contributions to associations, or to other forms of cooperation, within the basic structure. As we have seen, these contributions are estimated on the basis of particular configurations of contingencies, which are influenced in part by individual efforts and achievements, in part by social accident and happenstance. Contributions can only be locally defined as contributions to this or that association in this or that situation. Such contributions reflect an individual's worth (marginal usefulness) to some particular group. These contributions are not to be mistaken for contributions to society itself, or for the worth to society of its members as citizens. The sum of an individual's entitlements, or even of their uncompensated contributions to associations within society, is not to be regarded as a contribution to society. To this kind of contribution we can give no meaning; there is no clear or useful notion of an individual's contribution to society that parallels the idea of individual contributions to associations within society. Insofar as we compare the worth of citizens at all, their worth in a well-ordered society is always equal; and this equality is reflected in the system of basic equal liberties and fair opportunities, and in the operations of the difference principle.

Third and last, recall that in a Kantian view the parties are regarded as free and equal moral persons. Now freedom means a certain form of social institutions, namely, a certain pattern of rights and liberties; and equality in turn means, for example, that certain basic liberties and

opportunities are equal and that social and economic inequalities are regulated by principles suitably expressive of equality. Moral persons are those with a conception of the good (a system of final ends) and a capacity to understand a conception of justice and to follow it in their life. Of course, we cannot define free and equal moral persons as those whose social relations answer to precisely the principles that are agreed to in the social contract. For then we should have no argument for these principles. But it is no accident that once the parties are described in terms that require some social expression, the first principles of justice are themselves institutional and apply to the public structure of society. The content of the two principles fulfills this expectation. And this is in contrast, for example, with utilitarianism which takes as basic the capacity for pleasure and pain, or for certain valuable experiences. Nevertheless, the social manner in which the parties are described does not mean a lapse into some kind of holism; what results is a conception of a well-ordered society regulated by the two principles of justice.

VII

Now I come to the last point: namely, why it is that, although society may reasonably rely on a large element of pure procedural justice in determining distributive shares, a conception of justice must incorporate an ideal form for the basic structure in the light of which the accumulated results of on-going social processes are to be limited and corrected.

First a remark about pure procedural justice: the two principles make considerable use of this notion. They apply to the basic structure and its system for acquiring entitlements; within appropriate limits, whatever distributive shares result are just. A fair distribution can be arrived at only by the actual working of a fair social process over time in the course of which, in accordance with publicly announced rules, entitlements are earned and honored. These features define pure procedural justice. Therefore, if it is asked in the abstract whether one distribution of a given stock of things to definite individuals with known desires and preferences is more just than another, then there is simply no answer to this question.³

Thus the principles of justice, in particular the difference principle, apply to the main public

³ On pure procedural justice, see *ibid.*, pp. 64, 66, 72ff, 79, 84-89, 274-280, 305-315.

principles and policies that regulate social and economic inequalities. They are used to adjust the system of entitlements and earnings and to balance the familiar everyday standards and precepts which this system employs. The difference principle holds, for example, for income and property taxation, for fiscal and economic policy. It applies to the announced system of public law and statutes and not to particular transactions or distributions, nor to the decisions of individuals and associations, but rather to the institutional background against which these take place. There are no unannounced and unpredictable interferences with citizens' expectations and acquisitions. Entitlements are earned and honored as the public system of rules declares. Taxes and restrictions are all in principle foreseeable, and holdings are acquired on the known condition that certain corrections will be made. The objection that the difference principle enjoins continuous and capricious interference with private transactions is based on a misunderstanding.

Again, the two principles of justice do not insist that the actual distribution reflect any observable pattern, say equality, nor any measure computed from the distribution, such as a certain Gini coefficient (as a measure of the degree of equality). What is enjoined is that (permissible) inequalities make a certain functional contribution over time to the expectations of the least favored. The aim, however, is not to eliminate the various contingencies from social life, for some such contingencies seem inevitable. Thus even if an equal distribution of natural assets seemed more in keeping with the equality of free persons, the question of redistributing these assets (were this conceivable) does not arise, since it is incompatible with the integrity of the person. Nor need we make any specific assumptions about how great these natural variations are; we only suppose that, as realized in later life, they are influenced by many kinds of contingencies. Institutions must organize social cooperation so that they encourage constructive efforts. We have a right to our natural abilities and a right to whatever we become entitled to by taking part in a fair social process. The two principles of justice define the relevant fair process and so whatever distributive shares result are fair.

At the same time, these principles specify an ideal form for the basic structure in the light of which pure procedural processes are constrained and corrected. Among these constraints are the

limits on the accumulation of property (especially if private property in productive assets exists) that derive from the requirements of the fair value of political liberty and fair equality of opportunity, and the limits based on considerations of stability and excusable envy, both of which are connected to the essential primary good of self-respect.⁴ We need such an ideal to guide the corrections necessary to preserve background justice. As we have seen, even if everyone acts fairly as defined by rules that it is both reasonable and practicable to impose on individuals, the upshot of many separate transactions will undermine background justice. This is obvious once we view society, as we must, as involving cooperation over generations. Thus even in a well-ordered society, adjustments in the basic structure are always necessary. What we have, in effect, is an institutional division of labor between the basic structure and rules applying directly to particular transactions. Individuals and associations are left free to advance their ends more effectively within the framework of the basic structure secure in the knowledge that elsewhere in the social system the necessary corrections to preserve background justice are being made.

The essential point, then, is that the need for a structural ideal to specify constraints and to guide corrections does not depend upon injustice. Even with strict compliance with all reasonable and practical rules, such adjustments are continually required. The fact that actual political and social life is often pervaded by much injustice merely underlines this necessity. A procedural theory that contains no structural principles for a just social order would be of no use in our world, where the political goal is to eliminate injustice and steer change towards a fair basic structure. The notion of a well-ordered society provides the requisite structural principles and specifies the overall direction of political action. There is no rational basis for preventing or eliminating injustice if such an ideal form for background institutions is rejected.

VIII

I conclude with some general remarks. I have tried to show that once we think of the members of society as free and equal moral persons, it is natural for a social contract view to take the basic structure of society as the primary subject of justice. Yet in view of the distinctive features of this structure, the

⁴ See *ibid.*, pp. 224-227, 277f; 534-537, 543-546.

initial agreement must be characterized in such a way that: (a) it brings out the difference between this agreement and all other agreements; and (b) it enables us to take account of the fully social nature of human relationships. Throughout my aim has been to set the stage for a reply to idealism.

The problem is this: to develop a viable Kantian conception of justice the force and content of Kant's doctrine must be detached from its background in transcendental idealism and given a procedural interpretation by means of the construction of the original position. It must then be shown that this construction, which draws upon the idea of the social contract, is not subject to the cogent objections that idealists raised against the contract doctrine of their day. Thus Hegel thought that this doctrine confused society and the state with an association of private persons; that it permitted the

general form and content of public law to be determined by the contingent and specific private interests of individuals; that it could find no place for the fact that it is not optional (up to us) whether or not we belong to society. For Hegel the social contract was an illegitimate and unwitting extension of ideas at home in and limited to (what he called) civil society.⁵

I have tried to show how the conception of justice as fairness avoids these and similar objections. The procedural interpretation of Kant's view not only satisfies the canons of a reasonable empiricism, but its use of the idea of the social contract meets Hegel's criticisms. At the same time, since it proceeds from a suitably individualistic basis, it presents the details of a moral conception that can take appropriate account of social values without falling into organicism.⁶

Harvard University

Received September 1, 1976

⁵ See *Philosophy of Right*, tr. by T. M. Knox (Oxford, 1942), pp. 58f, 156f, 186.

⁶ I am indebted to Joshua Cohen, Joshua Rabinowitz, and T. M. Scanlon for valuable discussions on the topic of this paper. In Sections 5 and 6 I indicate the lines of a reply to David Gauthier's illuminating critique "Justice and Natural Endowment," *Social Theory and Practice*, vol. 3 (1974), pp. 3-26. I regret that lack of space forbids a fuller exposition. I am grateful to Burton Dreben for many improvements.

X. THOUGHT AND ITS REFERENCE

RODERICK M. CHISHOLM

"What makes my idea of him an idea of *him*?"*

I

GIVEN the concept of a propositional attitude—for example, *acceptance*, *entertainment*, or *endeavor*—we have all that is needed to add a theory of objective reference, or intentionality, to logic and ontology. To make the theory adequate to the reference that is characteristic of imagination, perception, and self-awareness, we need only add the concept expressed by "being appeared to in a certain way" together with the various epistemic concepts definable in terms of "*h* is epistemically preferable to *i* for *S* at *t*." But we need not appeal to any "dimension of intentionality" other than what is involved in a propositional attitude. And we do not need to make reference to language, or to speech acts, or to words, or to any concept that is essentially linguistic.

II

We may define *proposition* intentionally, by reference to the concepts of acceptance and of possibility *de re*. Let us say:

D1 *p* is a proposition =Df. It is possible that there is someone who accepts *p*

Propositions, so conceived, are abstract objects, existing in every possible world.¹

We may say of a proposition that it entails certain other *propositions*; we may also say of a proposition that it entails certain *properties*. These two senses of entailment are distinguishable as follows:

D2 *p* entails *q* =Df. *p* is necessarily such that (a) if it is true then *q* is true and (b) whoever accepts it accepts *q*

The second clause insures that entailment, so con-

ceived, is much more restrictive than logical implication.

D3 *p* entails the property of being *F* =Df. *p* entails a proposition which is necessarily such that it is true if and only if something has the property *F*.

For example, the proposition *some men being mortal* entails the properties of *being a man* and *being mortal*. (I assume there is no property corresponding to "either being a man and a thief or being a man and a non-thief." But *all men being mortal* does not entail such properties.

The following concept, I suggest, is basic to the theory of intentionality, or objective reference. All types of objective reference may be explicated in terms of this concept and that of a propositional attitude.

D4 *p* implies *x* to have the property *F* =Df. There is a property *G* such that (i) only one thing can have *G* at a time, (ii) *p* entails the conjunction of *G* and the property *F*, and (iii) *x* has *G*

Now we may say that a proposition *p* *pertains to*—or *is about*—a particular thing *x*, provided only that *p* implies *x* to have some property.

III

The following definition suggests that propositions are the primary objects of belief.

D5 *S* believes that *h* =Df. The proposition that *h* is accepted by *S*

But beliefs and other propositional attitudes may also be *de re*. Not only may one accept the proposition that the next President is wise; one may also

* Ludwig Wittgenstein, *Philosophical Investigations* (Oxford, 1953), p. 177. (The published English version translates "*Vorstellung*" as "image" instead of "idea.")

¹ This intentional definition of *proposition* is essentially the same as the definition of *proposition* suggested by Leibniz and that of *thought* [*Gedanke*] suggested by Frege. As an alternative to "proposition," we could use "state of affairs." Then "is true" in subsequent definitions could be replaced by "occurs, takes place, or obtains." A number of the concepts here defined are discussed in further detail in my book *Person and Object: A Metaphysical Study* (London and La Salle, Ill., 1976).

believe with respect to the next President that *he* is wise. In such a case, one is related, by belief, not only to a proposition, but also to that individual thing which is the next President.

In recent years, *de re* belief has been exhibited as a complex linguistic phenomenon. It is assumed to involve a relation to certain sentences; the sentences are said to contain a special type of name (e.g., a "vivid name") involving complex causal relations between the individual thing thus named and the one who has the *de re* belief about it.² Yet all this seems to be entirely gratuitous, for *de re* belief need not involve such linguistic phenomena at all. Since propositions may themselves pertain to, or be about, certain things, we may say that, if one accepts a proposition pertaining to a certain thing *x*, then one's *belief* pertains to that thing *x*.

D6 *x* is believed by *S* to have the property *F* = Df.

S accepts a proposition which implies *x* to have the property *F*

Such expressions as "*x* is believed by *S* to be green" may be used to abbreviate such expressions as "*x* is believed by *S* to have the property of being green."³

IV

We next consider the concept of perception. I suggest that the "objective reference" that perception involves does not differ in principle from what we have already considered. Perception differs from non-perceptual belief sensuously and epistemically: it is accompanied by a way of appearing which may be said to justify it. But it does not involve any additional type of reference or "dimension of intentionality." Let us say:

D7 *S* perceives that *h* = Df. (1) *S* believes that *h*;

(2) *S* is appeared to in a way which is such that, being appeared to in that way tends to

confirm that *h*; and (3) it is evident for *S* that *h*.

If we wish to interpret "perceive" in such a way that the expression "veridically perceive" would be redundant, we could add a fourth clause to our definition, consisting simply of '*h*'. The epistemic locations in the second and third clauses of our definiens may be explicated by reference to epistemic preferability and belief.⁴

The second clause in the definiens insures that the sense of perception here defined is a somewhat rigid one (perhaps that suggested by "But did you really *perceive* that *h*?"'). Other, less rigid senses could readily be characterized in terms of it.

We now define the sense in which perception may be directed upon non-propositional objects:

D8 *S* perceives *x* to have the property *F* = Df.

There is a proposition *p* which is such that: *S* perceives *p*; and *p* implies *x* to have the property *F*.

(Note that the sentence "*S* perceives the man on the corner to be talking" is not referentially opaque if interpreted in accordance with D8. Taken together with "The man on the corner is the bank-robber," it will imply "*S* perceives the bank-robber to be talking.") We may say, finally:

D9 *S* perceives *x* = Df. There is a property such that *S* perceives *x* to have that property.

V

The reference to oneself that is involved in self-awareness implies a kind of "privileged access." But the reference is propositional and does not differ in principle from what we have considered. The "privileged access" is a function only of the epistemic status of the propositions involved.

² Compare David Kaplan, "Quantifying In" in *Words and Objections: Essays on the Work of W. V. Quine*, ed. by Donald Davidson and Jaakko Hintikka (Dordrecht, 1969), pp. 206-242; Ernest Sosa, "Propositional Attitudes *De Dicto* and *De Re*," *The Journal of Philosophy*, vol. 67 (1970), pp. 883-896; and Mark Pastin, "About *De Re* Belief," *Philosophy and Phenomenological Research*, vol. 34 (1974), pp. 569-575. Hintikka defends the view that a person *S* has a belief, with respect to a certain thing *x*, only if *S* knows, with respect to *x*, that *x* is identical with *x*; but he does not attempt to explicate the latter *de re* epistemic location. See Jaakko Hintikka, *Knowledge and Belief* (Ithaca, N.Y., 1962), Chapter 6.

³ Some may feel that, if one's belief is thus to be about a certain thing, one should be somewhat closer, epistemically, to the thing in question. Thus we could replace the definiens of D6 by: "There is a property *C* such that (i) only one thing can have *C* at a time, (ii) *S* knows a proposition implying *x* to have *C*, and (iii) *S* accepts a proposition which implies *x* to have the property *C* and the property of being *F*." I assume that knowledge may be defined in terms of acceptance and our undefined epistemic location. I have discussed these questions in more detail in "Knowledge and Belief: '*De Dicto*' and '*De Re*,'" *Philosophical Studies*, vol. 29 (1976), pp. 1-20.

⁴ Thus the absolute locution "*p* tends to confirm *q*" may be defined as: "Necessarily, for every *S*, if *p* is evident for *S*, and everything that is evident for *S* is entailed by *p*, then *q* has some presumption in its favor for *S*. Definitions of "certain," "evident," "acceptable," and "has some presumption in its favor" may be found in my book *Theory of Knowledge*, Second Edition (Englewood Cliffs, N.J., 1976), Chapter One.

Every person, to the extent that he knows anything at all, is given to himself in the following way. There are propositions that "present themselves" to that person: the propositions imply that *he* exists; they are true; and they are necessarily such that, if they are true; then they are certain for him. The propositions expressed by "I seem to see people" and "I feel depressed" could be examples; so, too, for certain propositions about the ways I am being appeared to. Let us say:

D10 *p* is such that it is self-presenting to *S* at *t* = Df. *p* is true at *t* and is necessarily such that whenever it is true then it is certain for *S*.⁵

In English such self-presenting propositions are most naturally expressed in sentences that use the first-person pronoun. It can be shown, I think, that the most plausible theory of the use of the first person pronoun is the following. Each person who uses the first person pronoun uses it to refer to himself and in such a way that, in that use, its *Bedeutung* or reference is himself and its *Sinn* or intention is his own individual essence.⁶ The latter concept may be defined as follows:

D11 *C* is an individual essence = Df. *C* is a property which is such that, for every *x*, *x* has *C* if and only if: *x* is necessarily such that it has *C*, and it is impossible that there is a *y* other than *x* such that *y* has *C*.

If what I have said is correct, the self-presenting propositions referred to above will be propositions implying one's own individual essence. And the individual essence of *no other* contingent individual thing is thus implied by anything that is certain for him. (Indeed, it may be doubted whether most of us know *any* propositions implying the individual essence of any other contingent thing.)

Castañeda has shown that a statement of self-attribution, such as "Jones believes that he himself is wise" (*S*), is not implied by the corresponding quantified statement, "There is an *x* such that *x* is identical with Jones and *x* is believed by *x* to be wise" (*Q*). Thus (*Q*) might be true and (*S*) false if: Jones reads the lines on his hand and takes them

to be a sign of wisdom; he doesn't realize the hand is his (for it is one of many protruding from a blanket); and he is unduly modest and entirely without conceit.⁷

What, then, does (*S*) tell us that (*Q*) does not? I suggest it tells us this: "Jones has an individual essence *H*; he accepts a proposition which is necessarily such that it is true if and only if whatever has *H* is wise." Self-attribution, so conceived, could thus be defined as follows in the terms we have been using:

D12 *S* believes himself to be *F* = Df. There is a *C* such that *C* is an individual essence of *S*; and *S* accepts a proposition which entails the conjunction of *C* and the property of being *F*.

Given this definition, we may say, of '*S* believes himself to be *F*' and "There is an *x* such that *x* is identical with *S* and *x* is believed by *x* to be *F*," that the first implies the second and not conversely. The definition presupposes (a) that, if *S* knows anything, then there are propositions entailing his own individual essence which make it certain for him that he exists, and (b) that he does not have such knowledge of propositions entailing the individual essence of any other contingent individual thing.

VI

Now we are in a position to consider the nature of *thinking*. Here we must refer to another propositional attitude—that of entertainment. The concept that may be expressed by "*p* is entertained by *S*," like that expressed by "*p* is accepted by *S*," concerns a relation between a person and a proposition. It may be called the basic concept of the theory of thinking; for, it would seem, thinking cannot be analyzed without presupposing the concept of entertainment.⁸

The logic of entertainment differs from that of acceptance or belief in several ways. Of the following two principles, the first but not the second has its analogue in the logic of belief: (E1) "Whoever entertains (*p* & *q*) entertains *p*"; and (E2) "Whoever

⁵ This definition presupposes that one and the same proposition may have different truth-values at different times. Hence some might prefer to use "state of affairs" where I have used "proposition" and to use "obtains" or "occurs" where I have used "is true."

⁶ This general view is defended in detail in the first chapter of *Person and Object*, *op. cit.*

⁷ Compare Hector-Neri Castañeda, "He': A Study in the Logic of Self-Consciousness," *Ratio*, vol. 7 (1966), pp. 130-157, and "On the Phenomeno-Logic of the I," *Akten des XIV Internationalen Kongresses für Philosophie*, Vol. III (University of Vienna, 1969), pp. 260-266.

⁸ Compare H. H. Price, *Belief* (London, 1969), p. 192.

entertains ($p \vee q$) entertains p ." (I presuppose that the proposition that would be expressed by a locution of the form " $p \& q$ " would be the same as that expressed by " $q \& p$ "; and similarly for " $p \vee q$ " and " $q \vee p$ ".) The following schematic principle also has its analogue in the logic of belief: (E₃) "Whoever entertains the proposition that it is *true* that h entertains the proposition that h ." And similarly if "true" is replaced by: "necessary"; "possible"; "desirable"; "false"; "impossible." But, it would seem, we should *not* affirm: "Whoever entertains the proposition that h entertains the proposition that it is false that h "; "Whoever entertains p entertains $p \vee q$ "; "Whoever entertains the proposition that it is impossible that h entertains the proposition that it is necessary that it is false that h ."

One may distinguish an "occurrent" from a "nonoccurrent" sense of belief or acceptance. But one may not similarly distinguish an "occurrent" from a "nonoccurrent" sense of *entertainment*. Entertainment, we could say, is necessarily occurrent. (Hence the number of propositions one can entertain at any time is relatively small.) And the distinction between the occurrent and non-occurrent sense of belief or acceptance would seem to presuppose the concept of entertainment. Thus we could say:

- D13 S believes in the occurrent sense at t that h = Df. The proposition that h is accepted by S at t ; and the proposition that he believes that h is entertained by S at t .⁹

VII

One sense of " x is thought about by S " is that expressed by " x is entertained by S ." A second sense of " x is thought about by S " might be expressed by saying: " S entertains a proposition about x ." The following definition is intended to capture both of these senses:

- D14 S has a thought that is directed upon x = Df. S entertains a proposition which implies x to have some property, or S entertains x .

Let us now add this definition:

- D15 x is thought about by S as having the property F = Df. S entertains a proposition which implies x to have the property F .¹⁰

Plato suggests in the *Parmenides* (Hypothesis V) that we cannot think of anything without thinking of it as existing. Our definitions enable us to see in what sense Plato's suggestion is true: we cannot think of anything as an F without entertaining a proposition that implies that there is an F .

Now we may note S 's thoughts may be directed upon a state of affairs or proposition p without it being the case that S is entertaining p .

The importance of this last point may be seen by considering the argument: "(1) S has never thought about flags flying on the Statue of Liberty. (2) S has thought about the President's favorite state of affairs being a strange one. And (3) the President's favorite state of affairs is flags flying on the Statue of Liberty. Hence (4) the President's favorite state of affairs is something that S has thought about. And therefore (5) that state of affairs which is flags flying on the Statue of Liberty is something that S has thought about." The premisses are consistent with each other, but the conclusion contradicts the first premiss. How can this be?

The first two premisses pertain to the *entertainment* of states of affairs; (1) tells us that S has never entertained that state of affairs which is flags flying on the Statue of Liberty, and (2) tells us that S has entertained that state of affairs which is the President's favorite state of affairs being a strange one. But proposition (5) does not entail the contradiction of what is intended by (1). For although (5) tells us that S 's thought is *directed upon* that proposition or state of affairs which is flags flying on the Statue of Liberty, it does not tell us that he is *entertaining* that proposition.

VIII

Thinking may be accompanied by imagery—by sensing or being appeared to. And we sometimes picture the states of affairs we are entertaining.

⁹ What is the relation between *entertaining* a proposition p and *deliberating* about the truth of a proposition p ? One may entertain p without deliberating about p . And one may deliberate about p without entertaining p . During the course of any period of time through which one deliberates about p there will be times at which one entertains p . But during such deliberation there may be moments during which one does not entertain p . Perhaps we could define " S deliberates about the truth of p " by saying: "There are certain propositions which S is entertaining, and he is doing so for the purpose of bringing it about that he accept p if and only if p is true."

¹⁰ We noted above (footnote 4) a way in which our definition of *de re* belief (D6) could be made more restrictive. Our definitions of *de re* thought (D14 and D15) could be modified in analogous ways.

("Picture" should here be taken in a broad sense and not restricted to visual imagery.) What is it, then, to picture something?

One might consider defining "*S* pictures an *F*" by saying "*S* is appeared to in a certain way and is entertaining the proposition that in being thus appeared to he perceives something to be *F*."¹¹ But this would not give us a sufficient condition of what it is to picture an *F*. I could entertain the proposition that in sensing as I do now I'm perceiving something to be an elephant; yet I wouldn't thereby be picturing an elephant.

I suggest that we may say this:

D16 *S* pictures something to have the property *F* = Df. There is a way of being appeared to such that being appeared to in that way tends to confirm the proposition that something has the property *F*; *S* is appeared to in that way; and *S* is entertaining a proposition implying that in being appeared to in that way he is perceiving something to have *F*.¹²

In place of "*S* pictures something to have the property of being a so-and-so," we could also say, more simply, "*S* pictures a so-and-so." The definition may be adapted in an obvious way to the locution "*S* pictures just *n* things as having *F*." And "*S* pictures *x*" could be spelled out after the manner of "*S* perceives *x*," in D9 above.¹³

IX

We are now in a position to answer the philosophical question with which we began: "What makes my idea of him an idea of *him*?"

The answer is, simply: "To say that I now have an idea that is an idea of *him* is to say that I now accept or am entertaining a proposition about him"

But the plausibility of any philosophical theory is a function of the extent to which the theory enables us to deal with philosophical puzzlement. In what follows, then, let us apply what we have said to three additional philosophical puzzles.

X

"Our thoughts, surely, constitute a subject-matter with respect to which we can obtain complete certainty. I can hardly be mistaken with respect to what it is that I am thinking, even though what it is that I am thinking may itself be mistaken. But yesterday I thought I was thinking about Albuquerque. And now I realize it was Tucumcari. How can this be?"

The solution is this. Yesterday I was entertaining a certain proposition *p* and I was certain that I was entertaining *p*; *p* implied something with respect to Tucumcari but nothing with respect to Albuquerque; and I thought that it implied something with respect to Albuquerque. From the facts (a) that there is a proposition which is such that it is certain to me that I am entertaining that proposition and (b) that proposition implies something with respect to a certain entity *x*, and is thus about *x*, it does not follow that it is certain to me that I am thinking about *x*.

XI

Our second puzzle concerns the status of certain nonexistent *intentionalia* or objects of thought.

The following seemed obvious to Meinong: If I happen to think of a golden mountain while not thinking of any other mountain, then the mountain I am thinking of can be said to be golden. The mountain I am thinking of can also be said to be a mountain, to be something that I am thinking of, and to be nonexistent. Again, if I happen to think of a round square, then one of the objects of my thought is both round and square and therefore, according to Meinong, it is not only nonexistent but also impossible. In such cases, he held, we have certain truths about objects which do not exist and, indeed, about certain objects that cannot exist.¹⁴ But how can a truth be about an object that does not exist?

It is often assumed that Russell's theory of descriptions shows us how to deal with Meinong's problem. But it is no disparagement of Russell's great contribution to philosophy to point out that

¹¹ Compare J. E. M. McTaggart, *The Nature of Existence*, Vol. II (Cambridge, 1927), pp. 107-108.

¹² It should be noted that, for purposes of brevity, certain locutions are introduced in this definition that have not previously been defined. But the definiens may be spelled out, in obvious ways, without these abbreviations.

¹³ I believe it should now be obvious that what we have said about perception, self-awareness, and picturing may readily be adapted to the type of reference that is involved in *memory*.

¹⁴ See Alexius Meinong, "The Theory of Objects" in R. M. Chisholm (ed.) *Realism and the Background of Phenomenology* (Glencoe, Ill., 1960), pp. 76-118.

it does not apply to "The mountain I am thinking of is golden." For consider what the paraphrase would be if we were to apply Russell's theory to Meinong's sentence. We would be saying: "There exists an x such that (i) x is a mountain I am thinking of, (ii) x is golden, and (iii) for every y , if y is a mountain I am thinking of, then y is identical with x ." This sentence, obviously, is false. For there are no golden mountains. But our original sentence—"The mountain I am thinking of is golden"—is true. And therefore the paraphrase is inadequate; for if a sentence is true, then no adequate paraphrase of it will be false.

What shall we say, then, of Meinong's statement "The mountain S is thinking of is golden"? We might try to formulate the truth that is here involved by reference to *entertainment*: "The propositions that S is entertaining entail the proposition that there is a golden mountain, and they do not entail the proposition that there is more than one golden mountain."

It may be, however, that when we use the expression, "The mountain S is thinking of," we could also express ourselves by saying "The mountain S is picturing to himself." In this case, it may be that, although S is entertaining propositions which entail that there are many golden mountains, we could yet say, "The mountain S is picturing is golden." We may understand this by

adapting D16 above to " S pictures just one thing as being a golden mountain."

XII

"(1) In order for there to be reference, there must be an intention to communicate, the performance of a speech-act, and hence the existence of some kind of medium of communication, as well as some uptake on the part of a receiver. But (2) thinking about individuals can take place without such linguistic happenings. Therefore (3) there is no reference in thought. But if there is no reference in thought how can one be said to think *about* a certain particular thing?"¹⁵

To deal with the puzzle, we need not imagine, as some philosophers have supposed, that thinking always does involve some kind of linguistic activity such as the inner performance of speech acts. We may simply appeal to D14, our definition of what it is to have a thought that is directed upon a certain thing x . *That* tells us how it is that one can be said to think *about* a certain particular thing.¹⁶ Hence the first premise in the formulation of the puzzle is mistaken. It is *not* the case that, in order for there to be reference there must be speech-acts or other such linguistic events. Linguistic activity presupposes thought, but not conversely.

Brown University

Received September 1, 1976

¹⁵ This puzzle is suggested, though not quite as I have put it, by Zeno Vendler in "Thinking of Individuals," *Nous*, vol. 10 (1976), 35-46; see especially, pp. 36-37.

¹⁶ And so we might conclude, as Vendler does, by asking: "Did we not learn in school that speech is nothing but the expression of thought?" (p. 45) But contrast A. J. Ayer, *Thinking and Meaning* (London, 1947); Wilfrid Sellars, "Empiricism and the Philosophy of Mind," *Minnesota Studies in the Philosophy of Science*, vol. I (1956), pp. 253-329; and P. T. Geach, *Mental Acts* (London, 1957). Two fundamentally different points of view on this topic are expressed in the "Chisholm-Sellars Correspondence on Intentionality," *Minnesota Studies in the Philosophy of Science*, Vol. II (1958), pp. 521-539.

XI. HOW TO USE PROPOSITIONS

GILBERT HARMAN

WHAT makes Mary's thought a thought about John? Let us consider the following answer. "Mary's thought consists in the entertaining of a proposition, which on this occasion happens to be a proposition about John." Whether this is a useful answer clearly depends on whether we can suppose the proposition Mary is entertaining serves as a real intermediary between Mary's thought and John. The answer is therefore useful only if coupled with a conception of propositions that allows us to divide our original question into two nontrivial further questions: (1) What is it for Mary to entertain a given proposition? and (2) What is it for a given proposition to be a proposition about John?

Suppose, for example, we take a proposition to be a complex of the things it is about, so that the proposition *John is wise* is a complex whose elements include John and (perhaps) the attribute of wisdom. Then question (2) is easily answered: a given proposition is a proposition about John if and only if it contains John as one of its elements. But we are no closer to answering our original question, "What makes Mary's thought a thought about John?" For we have merely replaced that question with the equally difficult question, "What is it for Mary to entertain a proposition in this sense that contains John as a constituent?" Given this conception of a proposition, a proposition does not serve in any useful way as an intermediary between Mary's thought and its object, John.

If propositions are to have a real function in this context, they must be things Mary uses in her thinking. This suggests that propositions must be mental entities. On the other hand, propositions also seem to be abstract entities, universals rather than particulars. Mary can have the thought that John is wise on several different occasions, and other people can also have this thought at various times. The same proposition is presumably involved in these various instances; in that sense the proposition is a universal rather than a particular. But how can Mary and other people use an abstract entity of this sort in their thinking? Or, to put the question differently, how can Mary use the same mental

entity at several different times and how can others use the same mental entity Mary uses?

One way to resolve this issue is to distinguish proposition types from proposition tokens, just as we distinguish sentence types from sentence tokens. Whenever Mary entertains the proposition *John is wise*, she entertains a mental instance or token of this proposition. She and other people entertain the same proposition on various occasions by entertaining various mental instances of the same proposition type.

Now, indefinitely many propositions are available to Mary for use in her thinking. It is hard to see how this could be so unless these propositions were composed from a finite set of concepts and ideas, just as indefinitely many sentences are constructed from a finite set of words. When Mary entertains a proposition, she forms a mental instance of the relevant type. But we need not ascribe to Mary an indefinite number of primitive abilities to form instances of each proposition she is capable of entertaining. Instead we can ascribe to her a finite number of primitive abilities to form instances of ideas or concepts and an ability to combine these ideas or concepts together in a proposition.

There are, then, certain reasons—I am not saying they are absolutely compelling—for seeing two similarities between propositions and sentences: the type-token distinction is relevant in both cases and both propositions and sentences are constructed from a finite stock of elements. An obvious third similarity is that "aboutness" is involved in both cases—Mary uses propositions in having thoughts about John and uses sentences in making remarks about him. We might express these similarities by saying that we have two kinds of "symbol system" here, or two kinds of "system of representation". Or we might speak of the "language of thought", being careful of course not to imply that possession of language, in the ordinary sense of language, is necessary for thought. In any case we must acknowledge differences between communication and thought.

Now, in order to say what it is for a proposition to be about John, we must make further assump-

tions about the system of representation Mary uses in her thinking. So let us suppose that among the elements of her system are a number of what we will call mental names. If Mary believes that there is a unique thing satisfying certain conditions C_1 , C_2 , C_3 , she can introduce a new mental name a into her system by forming the beliefs that a is C_1 , that a is C_2 , and that a is C_3 . This name functions as a name of the unique thing satisfying these conditions if there is one; otherwise it does not name anything. Moreover, the name continues to name this thing, as long as Mary uses it, even if nothing or some different thing should be the unique thing satisfying those of her beliefs involving the name a .

In this view Mary can believe about John that he is wise without being able to identify John in any nonquestion-begging way. In other words, there does not have to be a nonquestion-begging condition C that John uniquely satisfies such that Mary believes that something uniquely satisfying C is wise. It is enough if Mary believes a proposition of the form " a is wise," using a mental name a as a name of John. When Mary first introduced that name, she had to be able to identify him in a nonquestion-begging way, but not at any later time. She will of course later be able to identify someone as a , but that question-begging identification cannot be used to determine what a names.

Let us suppose, furthermore, that one of Mary's mental names is a name of herself that functions as a first person pronoun. For example, Mary tends to treat beliefs and desires containing this name as more immediately relevant to her actions than beliefs and desires containing other names of her (which she may not realize are names of her). When Mary expresses in English propositions containing this name, she uses "I," "me," "my," or "mine." Propositions not containing this name but containing a different mental name of Mary are not appropriately expressed in English by her in these terms.

Now one way a proposition can be about John is by containing a mental name of John. There is also, let us suppose, another way. It can contain a mental image of John. Mental images are aspects of Mary's imaginative or perceptual experience, aspects that play a distinctive role in Mary's system of mental representation. When Mary sees John, she has subjective perceptual experience which represents John as several feet away, facing her, smiling, with right arm extended. In perceiving John in this way, Mary entertains the proposition that she sees John several feet away, facing her, smiling, with right

arm extended. This is to entertain a proposition about John; in perceiving John, Mary is thinking about John. Her thinking does not merely *accompany* her perceptual experience, however; she has the thought by having the experience. Her perceptual experience is part of a mental instance of the complex perceptual propositions she is entertaining. This instance of the proposition therefore consists partly of images, of which one is an image of John. That is why the proposition is a proposition about John.

This sort of representation by images is also involved when Mary imagines John or pictures him as several feet away, facing her, smiling, with right arm extended. Mary pictures John in order to think of him in this way. Her thought does not merely accompany her imagery. Her imagery is part of an instance of the proposition she is entertaining. She entertains the proposition by having the imagery. The (instance of the) proposition she entertains is (an instance of) a proposition about John because it contains a mental image of John.

What makes a mental image an image of John? For some cases a causal answer seems appropriate. A perceptual image is an image of John because the presence of John is in a familiar way causally responsible for important aspects of the image. A memory image is an image of John because of a causal connection between it and an earlier perceptual image of John that was causally dependent on John. The image Mary forms when she pictures John or when she wrongly thinks she sees John may in a different way be causally dependent on earlier perceptual images of John.

But there are also cases in which Mary pictures someone or something she has never perceived. These cases may involve a mixture of types of representation, mental images combined with mental names—labelled mental image. When Mary pictures John Hancock signing the Declaration of Independence, she forms an image of someone signing something. This is an image of John Hancock because it is labelled with a mental name of John Hancock. This name is a name of John Hancock because of the way Mary introduced it into her system of mental representation.

Here then is one way in which we might hope to use propositions in order to explain what makes Mary's thought a thought about John: Mary's thought is about John if she entertains a mental instance of a proposition containing a mental name or mental image of John.

Chisholm also hopes to use propositions for this

purpose but without making such strong assumptions about the structure of propositions, in particular without assuming that propositions can contain mental names or images. However, I do not see how this can be done.

His basic definition says, roughly, that a proposition p is about something x if and only if

- (1) p entails¹ that there is one and only one F and
- (2) x is the one and only F .

This is circular unless there is a restriction on what can replace ' F .' Otherwise, we could replace ' F ' with "thing identical with x " to get this:

- (1') p entails that there is one and only one thing identical with x
- (2') x is the one and only one thing identical with x .

Because of the position of the free variable ' x ' in (1'), (1') is equivalent to this:

- (1'') p entails a proposition saying about x that there is one and only one thing identical with it.

This is circular, like defining " p is about x " as " p entails a proposition that is about x ." If Chisholm's definition is to avoid this sort of circularity, he must not allow replacements of ' F ' that contain the free variable ' x .'

When such a restriction is placed on ' F ,' however, his account implies that Mary has a thought about John only if she can identify him in a non-question-begging way. But why should Mary have to be able to identify John in this way? Perhaps all she remembers of him is that he is wise. The account I sketched using mental names allows for this possibility, and Chisholm has given us no reason to rule it out.

There is a related objection to Chisholm's analysis of what it is for Mary to believe herself to be wise, a belief she would express in English as "I am wise." She does not have this belief if, reading the lines on a hand without realizing that it is her hand, she believes *this person is wise*. According to Chisholm, for Mary to have the relevant belief is for her to accept a proposition entailing that something C is wise, where the property of being C is an "individual essence" of Mary. But clearly Mary can accept such a proposition even in the wrong case, for she might accept a proposition

entailing that something identical with this person is wise, and the property of being identical with this person is an individual essence of Mary even though she is not aware of this. It might seem that this objection could be avoided by requiring that the property of being C be a purely qualitative individual essence, but then there is no reason to suppose that Mary has such an essence and there is certainly no reason to think that Mary accepts any propositions entailing that something is C , if the property of being C is a purely qualitative individual essence of Mary. The alternative account in terms of mental first person pronouns is not subject to this sort of objection.

A further objection is that Chisholm's account implies that Mary has a thought about John on certain occasions when she does not seem to be thinking of him. Suppose, for example, that she has the thought that one and only one student will be elected president of the student association next year and suppose that in fact John will be elected. Chisholm's account implies that Mary's thought is about John and indeed that she has the thought about John that he will be elected president of the student association next year. That is counter-intuitive.

Chisholm is aware of this objection and proposes to meet it by requiring Mary to *know* that one and only one student will be elected president of the student association next year. This has the curious result that Mary may entertain a proposition about John without her thought being about John. But anyway suppose Mary does know that one and only one student will be elected. It still seems incorrect to say she has the thought about John that he will be elected president of the student association next year.

According to the account I sketched using mental names, on the other hand, Mary's merely existential thought is not yet a thought about John, since she is not entertaining a proposition that contains a mental name of John. Knowing that one and only one student will be elected president of the student association, Mary could always introduce a new mental name for that person into her system of mental representation, if she had any reason to. She could then form the thought about that person, John, that he will be elected president of the student association next year. But normally there would be no point to this. Mary would nor-

¹ Entailment for Chisholm is a very strong relation between propositions; p entails q only if, necessarily, if p is true, q is, and anyone who believes p also believes q .

mally have to reason to introduce a new mental name in this context. So Mary would not normally form a thought about John in this context.

More needs to be said about this, of course. We need an account of the reasons for using mental names in a system of mental representation parallel to Strawson's account in *Subject and Predicate* of the reasons for using proper names in discourse. I leave this matter for future research.

I turn now to a different issue. Chisholm tries to explain what makes Mary's thought a thought about John without making any assumptions about the structure of propositions. But it is not clear that he succeeds. His basic definition as I have represented it seems to make use of the notion of a proposition's having existential form. Clause (1) reads, remember, " p entails that there is one and only one F ." Since entailment is a relation between propositions, we might suppose this means " p entails the proposition that there is one and only one F ," which assumes that we can speak of the structure of propositions.

In fact, however, Chisholm adopts the following interpretation: " p entails a proposition q which is such that, necessarily, q is true if and only if there is one and only one F ." On this interpretation of Chisholm's account, if p is true, Chisholm's special sort of entailment drops out of the picture in favor of necessary implication. For suppose that necessarily whenever p is true there is one and only one

F ; suppose p is true; and suppose x is the one and only one F . Then the truth of p is necessarily equivalent to there being one and only one thing such that both it is F and p is true. Since p entails itself, p entails something whose truth is necessarily equivalent to there being one and only one thing such that both it is F and p is true. So, by Chisholm's definition on this interpretation, p is about the one and only one thing such that both it is F and p is true; and x is that thing. So, if the truth of p necessarily implies that there is one and only one F , and if x is the one and only one F , and if p is true, p is about x , according to this interpretation. But this is unacceptable. Consider any necessarily existing entity that can be uniquely characterized as "the F ". Necessarily, if p is true, there is one and only one F . So any true proposition will be about any and every necessarily existent object that has a unique characterization, such as any of various numbers, attributes, and propositions. Indeed, a similar argument shows that p does not even have to be true. As long as p entails at least one true proposition, p will be about any and every uniquely specifiable number, proposition, and attribute, on this interpretation of Chisholm's account. I conclude, therefore, that this interpretation does not provide an adequate method for avoiding an appeal to the structure of propositions in the context of Chisholm's account.

Princeton University

Received August 1976

(Continued from outside back cover)

COMPLETE LIST OF MONOGRAPH SERIES

- No. 1 (1968) STUDIES IN MORAL PHILOSOPHY. Essays by David Braybrooke, Lawrence L. Haworth, G. P. Henderson, Jesse Kalin, Kai Nielsen, Jerome B. Schneewind, Michael Stocker.
- No. 2 (1968) STUDIES IN LOGICAL THEORY. Essays by James W. Cornman, Montgomery Furth, Alan Hausman and Charles Echelbarger, Jaakko Hintikka, Robert C. Stalnaker, Ted Honderich, Colwyn Williamson.
- No. 3 (1969) STUDIES IN THE PHILOSOPHY OF SCIENCE. Essays by Peter Achinstein, Joseph Agassi, Simon Blackburn, Mario Bunge, Bernard R. Grunstra, Keith Lehrer, D. H. Mellor, Lawrence Sklar, Michael Slote, Stephen Spielman.
- No. 4 (1970) STUDIES IN THE THEORY OF KNOWLEDGE. Essays by Norman Malcolm, W. Donald Oliver, Peter Unger, John L. Pollock, John Knox, Jr., Alan R. White, Frederick Stoutland.
- No. 5 (1971) DESCARTES' PHILOSOPHY OF NATURE. By James Collins.
- No. 6 (1972) STUDIES IN THE PHILOSOPHY OF MIND. Essays by Harold Brown, Haskell Fain and A. Phillips Griffiths, O. H. Green, Donald F. Henze, Moreland Perkins, Alan R. White.
- No. 7 (1973) STUDIES IN ETHICS. Essays by Charles V. Blatz, Claudia Card, Norman O. Dahl, R. L. Franklin, James D. Wallace, A. D. Woozley.
- No. 8 (1974) STUDIES IN MODALITY. By Nicholas Rescher with the collaboration of Ruth Manor, Arnold vander Nat and Zane Parks.
- No. 9 (1975) STUDIES IN EPISTEMOLOGY. Essays by Dennis M. Ahern, Robert Almeder, Jonathan Bennett, Fred R. Berger, Peter A. French, Margaret Gilbert, David K. Glidden, Mark Pastin, E. E. Sleinis.
- No. 10 (1976) NEGATION AND NON-BEING. By Richard M. Gale

PRINTED IN ENGLAND

by William Clowes & Sons, Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY
MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

AN INTRODUCTION
TO MODAL LOGIC

by E. J. LEMMON

In collaboration with DANA SCOTT

Edited by KRISTER SEGERBERG

Monograph No. 11

Oxford 1977

This is the latest title in this very impressive monograph series, and we shall be publishing it in March 1977. The text is a collaboration between two well-known logicians, who have produced a lucid introduction to one of the fastest growing branches of modern logic. Not only will the beginner find the book particularly helpful, but it also provides a synthesis of work in the field that will interest advanced students.

To enable you to judge for yourself the range and importance of this series, a complete list of the Monographs is given on the inside back cover. The last issue supplied to institutions on subscription was No. 7 (1973); as from 1977, the price of all monographs to institutions will be \$9.00 (£4.50 to UK subscribers, £5.40 to those overseas). We regret to announce that also from 1977 it will no longer be possible to supply the Monographs free of charge to individual subscribers, and that the cost to them will be \$4.50 (£2.25 to UK subscribers, £2.70 to those overseas).

(Continued on inside back cover)

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 14/NUMBER 3

JULY 1977

CONTENTS

- | | | | |
|--|-----|---|-----|
| I. PETER UNGER: <i>The Uniqueness in Causation</i> | 177 | V. ROBERT GRIMM: <i>Eventual Change and Action Identity</i> | 221 |
| II. ANTONY DUFF: <i>Psychopathy and Moral Understanding</i> | 189 | VI. RICHARD H. SCHLAGEL: <i>The Mind-Brain Identity Impasse</i> | 231 |
| III. HENRY E. KYBURG: <i>All Acceptable Generalizations are Analytic</i> | 201 | VII. ROBERT A. JAEGER: <i>Am I in the World?</i> | 239 |
| IV. JOSEPH RUNZO: <i>The Propositional Structure of Perception</i> | 211 | VIII. MICHAEL McDERMOTT: <i>Sets as Open Sentences</i> | 247 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes" (short or long). Moreover the editor will only consider papers whose authors certify that while under consideration with us, they will not be submitted elsewhere.

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 12,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The subscription for 1977 is \$28.00 for institutions, \$12.00 to individuals. Single issues will be \$8.00 to *all* purchasers. For Monograph prices, see below. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The complete series is listed on the back cover. In 1977 these will be available at \$9.00 to institutions and non-subscribers, \$4.50 to individual subscribers. *Please note that it is no longer possible to supply the monographs free of charge to individual subscribers.*



I. THE UNIQUENESS IN CAUSATION

PETER UNGER

THERE is a logical principle governing causation which, while it has consequences of considerable philosophic interest, never appears to be fully articulated or even recognized by any philosopher. I call it *the principle of the uniqueness of that which causes*. Expressed in terms of the verb "cause," and a notion of a *particular thing*, which delimits the relevant objects of the verb, the principle reads like this:

If some entity causes, or some entities cause, a particular thing, then nothing else causes that latter thing and no other entities cause it.

This is an entirely general statement concerning the logic of causation, though of course it does not imply that anything ever is caused. In the antecedent, I talk of entities instead of things. This is to make clear that the principle governs persons and other conscious beings, insofar as they may be causes, as well as events, facts, physical objects and other entities naturally called things. It is quite obvious, I suppose, that our principle dictates as follows: If John's putting poison in Jim's food caused a certain particular thing, say, Jim's death, then Bill's shooting Jim did not cause that latter thing. I want it to be just as clear that our principle dictates this as well: If John's putting poison in Jim's food caused Jim's death, then John himself, who is a man and not that first fact or event, did not cause Jim's death. For convenience, however, we may also put our principle wholly in terms of expressions for things: If something or some things cause a particular thing, then nothing else causes the latter (and no other things cause it).

What are we to think of this principle, this *uniqueness condition* with which it presents us? I will argue that, properly understood, the evidence conspires to support this condition. Further, it is only by acknowledging this condition that we may get the best account of "cause," both verb and noun, while demarcating the sense most directly relevant to the topic of causation. Finally, I shall indicate some of the consequences of this principle.

1. ON THE OBJECTS OF CAUSATION

Our principle concerns causation, and not any philosopher's invention however closely or interestingly related to it. This is due to our use of the verb "cause" in our statement of the principle. We are to test the principle by examining certain sentences which employ this verb. Avoiding some trivially irrelevant sentences with "cause" is trivially easy. Avoiding some other irrelevant sentences demands a bit more thought. This thought concerns the objects of causation. Our principle, we may say, only concerns those cases where the objects of causation are *particular things*, and where the object of the verb picks out a single particular thing.

Accordingly, the following sentence is never intended as an instance of our principle:

Drunken driving causes accidents.

For, the object word, "accidents," does not refer to any particular thing. Thus, our principle allows the following related sentence, denying uniqueness of what causes, to be consistent, as it obviously is:

Drunken driving causes accidents, and so do many other things.

The sentences relevant to testing have an object for the verb which refers to a single particular thing.

It is less obvious that the following sentences are also not directly relevant to the question of whether or not our principle is correct:

Bill's sneezing caused Betty to catch a cold.

Bill caused Betty to catch a cold.

But they are not directly relevant. The point is that Betty's catching a cold may occur a number of times, or obtain on a number of occasions. Bill may cause her to catch a cold one time, someone else or something else may cause her to another time. Accordingly, our principle allows the following sentence to be actually consistent:

Bill's sneezing caused Betty to catch a cold, and so did something else.

Bill caused Betty to catch a cold, and so did someone else.

Without the explanation just given, to the effect that no particular thing is picked out by the object, many readers, I submit, would hear these sentences as inconsistent. This is because they would be *assuming* that a particular thing *was* picked out as caused, a particular case of Betty's catching a cold, and then our principle would govern their thinking which proceeded from that assumption. In that way, our principle receives some support, I think, even where the conditions are not proper for a direct logical test of it.

We obtain proper conditions when we begin with such sentences as these, where a particular object is picked out:

Bill's sneezing caused Betty to catch a cold for the second time.

Bill caused Betty to catch a cold for the second time.

For now, only one particular thing can answer to the object description: presumably, the event of Betty's catching a cold for the second time; possibly, the fact of her doing so, which may be the same thing as that event. In any case, we now encounter these directly relevant sentences:

Bill's sneezing caused Betty to catch a cold for the second time, and so did something else.

Bill caused Betty to catch a cold for the second time, and so did someone else.

These sentences not only appear to be inconsistent; they actually are. Thus, the sentences just above imply a uniqueness for their subject in relation to the object. And, it is the verb "caused" which is responsible for this implication. For with other verbs, otherwise identical sentences exhibit no inconsistency, nor even any appearance of one:

Bill's sneezing helped Betty to catch a cold for the second time, and so did something else.

Bill observed Betty to catch a cold for the second time, and so did someone else.

This contrast confirms the idea that there is a uniqueness in causation, and that this uniqueness is not to be found just anywhere. For example, it does not obtain in the case of observation.

Even where an inconsistency is present, it may not

appear to be. I think it clear, for example, that the sentence:

Men aren't men anymore

is an inconsistent one. But hardly anyone would hear it as inconsistent. The hearer's mind would proceed forthwith to an idea likely for a speaker to convey by means of the sentence, though not actually asserted by it, for example, that men aren't as virile as they used to be. The context would determine, more or less precisely, the consistent idea to which the mind would proceed.

Various causal sentences will likewise fail to appear inconsistent, even when they actually are so. One example of this is the sentence:

The flame caused the pan to get hotter at that time, and so did something else, the hot coals.¹

I think that what this sentence actually expresses can be rendered as follows, so that the inconsistency emerges:

The flame caused an increase in the amount of heat in the pan at that time, and the hot coals also caused that increase.

The idea towards which the hearer's mind naturally proceeds, however, may be this rather different one:

The flame caused the pan to get hotter at that time, and the hot coals caused it to get still hotter (just afterwards).

This is of course consistent, and in no way violates any uniqueness condition. The flame causes one increase and the hot coals cause another. Those are two different things, and so they may be caused by things each quite distinct.

Another sort of example, where no comparative condition is explicitly involved, occurs in the realm of the mental:

Bill caused Betty to be upset then, and so did someone else, John.

Again, I think that what the sentence actually expresses is inconsistent. But the mind will be apt to handle the sentence differently. For one, it may well move toward an idea involving comparatives:

Bill caused Betty to be upset then, and someone else, John, caused her to be still more upset.

This latter sentence, concerning a different matter,

¹ The apparently problematic status of this sort of sentence, and of the "mentalistic" sort of sentence soon to follow, was suggested to me by William Ruddick.

is of course consistent. For another, the mind may move toward an idea involving different objects of the mental state or states:

Bill caused Betty to be upset at him then, and someone else, John, caused Betty to be upset at him.

This too is quite consistent. But here, each person is said to cause something different. What our original sentence actually asserts, however, is neither of these, nor anything of the like. Its content is rendered, perhaps more clearly, in this following sentence:

Bill caused Betty to be in a certain state then, that of being upset, and so did someone else, John.

This last sentence, which is indeed relevant, is quite clearly an inconsistent one.

While it is not to be found just anywhere, at the same time, this causal uniqueness is no narrow or special matter, confined to the verb "cause" and close cognates. First, general verbs of causality exhibit the same logical feature; we find it with "bring—about—" and "make—(happen) —." Thus, the sentences we should expect them from do give us the revealing inconsistencies with these verbs:

Bill's sneezing brought it about that Betty caught a cold for the second time, and so did something else.

Bill made Betty catch a cold for the second time, and so did someone else.

Verbs with a causal element, but also with other more specific implications also conspire to suggest our uniqueness.

The verb "move" may be used intransitively, requiring no object but only a subject, as in the sentence "The stick moved." Here, the verb has no implication of causation. Alternatively, the verb may be used transitively, where an object of the verb is at least implied. This occurs in the sentence, "The man moved the stick." In all such cases, there is an implication of causality, the motion of what is moved is caused. In this case, the man presumably caused the stick to move; if he did not, then presumably he at least caused something which in turn caused the motion in question, or whatever. In any case, where the object of "move" picks out a single particular thing, a uniqueness condition applies, quite in parallel with the one with "cause." The same applies with many other transitive

causal verbs: "create," "destroy," "build," "construct," "write," "compose," "paint," "open," "close," "fill," "empty," and so on, and so forth. To test directly, we must be sure the sentence picks out a single, particular thing as caused. For example, John may paint the fence, but also Bill may paint it another time. Here, John may be causing a particular coat of paint to go on the fence, or something of the like, and Bill is causing another to go on. Normally, when speaking of a fence being painted, we are not thinking of various coats, or various paintings, or whatever. Thus, we would normally hear the following as inconsistent:

John painted the fence, and so did someone else.

This appearance serves to confirm our uniqueness condition. At the same time, we now can see that such a sentence is not directly relevant for our logical test, and that we move toward such a test sentence with:

John painted the fence at that time, and so did someone else.

I am not saying that this latter sentence is logically adequate to our purpose. I am only suggesting that we have intuitions regarding a certain uniqueness in the verb "paint," and that this goes along with our intuition concerning the causal nature of the verb.

Let us suppose that John painted the front of the fence, and Bill the back; and that when they finished, the fence was painted. Did John paint the fence? I think not. John painted the front of the fence. Why do we have this feeling? Because John, we feel, caused the front of the fence to be painted, or to acquire a coat of paint, or whatever. But he did not cause the back to do so, and so he did not cause the fence to be painted. Rather, John and Bill caused the fence to acquire a coat of paint, or whatever; and so neither one, but the two of them, painted the fence.

Let us suppose that a monkey opens a door with a key. In casual talk, someone might say that the key opened the door. Someone else might say that the monkey did. If what they both say is really true, we should be able to conjoin them to form a truth:

The monkey opened the door then, and so did the key.

But this seems inconsistent, as does the logical consequence of it:

The monkey opened the door then, and so did something else.

As something inconsistent can't be taking place, what is actually going on? I think that the proper answer here is that keys don't really open doors; but that they are just used to open doors, by monkeys, people and so on, and perhaps are otherwise involved in the opening of various doors. In any case, however, the main point for us here is this: The inconsistency we feel in attributing the opening to each of two things at once goes along with our feeling that, in opening a door, something or someone causes some particular thing, and the thing caused involves or concerns the door. Perhaps, he or it causes the door to open, or to become open or whatever. Accordingly, as opening a door is causal, the uniqueness we feel present in cases of such openings confirms our idea about the uniqueness in causation.

I have been deliberately vague in my characterization of what is caused, calling it only a *particular thing*. My reason is that I am not quite sure as to the nature of these things, but wish to convey the idea of the relevant particularity. Many would urge that these particular things are events, particular events. If I insisted on particular events, an argument would be necessary. For, first, Roderick Chisholm has denied that events are particulars, claiming them always to be a certain sort of universal.² And, second, Zeno Vendler, while seeming to allow that events are particulars, argues that events are seldom if ever caused. He seems to think that what are caused are particular facts.³ Our statement in terms of particular things is intended to be neutral amongst these various positions. But unless we have the particularity on which I am focusing, we will have no coherent idea of *causation* at all, but only some philosopher's invention. For present purposes, then, I may assume that our notion of causing a *particular thing* is adequately clear.

Often, however, historical discussions of causation have not even been this clear. In his *System of Logic*, John Stuart Mill offers the following definition of "the Law of Causation":

² Roderick Chisholm, "Events and Propositions," *Nous*, vol. 4 (1970), pp. 15-24. Also, see Donald Davidson's reply to Chisholm, "Events as Particulars," *Nous*, vol. 4 (1970), pp. 25-32.

³ Zeno Vendler, "Effects, Results and Consequences" in R. J. Butler (ed.), *Analytical Philosophy* (Oxford, 1962). Also, see replies by S. Bromberger and W. H. Dray, and Vendler's reactions, all in the volume just cited.

⁴ John Stuart Mill, *A System of Logic* Book III, Chapter V, Section 2.

⁵ Bertrand Russell, "On the Notion of Cause," *Proceedings of the Aristotelian Society*, vol. XIII (1912-13), reprinted in *Mysticism and Logic* (London, 1932), where the quoted passage is on p. 186.

⁶ Mill, *op. cit.*, Book III, Chapter X, Section 1.

The Law of Causation, the recognition of which is the main pillar of inductive science, is but the familiar truth, that invariability of succession is found by observation to obtain between every fact in nature and some other fact.⁴

While Mill speaks of a fact as the thing caused, the most likely interpretation of his meaning is that it is a universal fact, which may be instanced on numerous occasions. Indeed, in his interpretation of Mill's meaning, Bertrand Russell speaks of an event as caused, and it is an event which is a universal.

An "event", in the statement of the law, is obviously intended to be something likely to recur, since otherwise the law becomes trivial. It follows that an "event" is not a particular, but some universal of which there may be many instances. . . . What is meant by an "event" is something like striking a match, or dropping a penny into the slot of an automatic machine.⁵

Now, if one thinks, as Mill seems to do, in terms of the causing of such universals, it is all too easy to miss out on our uniqueness condition, which applies only to the causing of particular things. This would be so even if, as Mill himself does, one had thoughts which were extremely congenial to our condition. Indeed, in his most pertinent chapter, "Of the Plurality of Causes: and the Intermixture of Effects," Mill is most congenial to our way of thinking:

. . . We have regarded *a b c d e*, the aggregate of the phenomena existing at any moment, as consisting of dissimilar facts, *a b c d* and *e*, for each of which one, and only one, cause needs be sought; the difficulty being only that of singling out this one cause from the multitude of antecedent circumstances, A, B, C, D, and E. The cause indeed may not be simple; it may consist of an assemblage of conditions from which the given effect could result.

If such were the fact, it would be comparatively an easy task to investigate the laws of nature. But . . . it is not true that the same phenomena is always produced by the same cause; the effect *a* may sometimes arise from A, sometimes from B.⁶

Here, Mill begins with what appears to be a discussion of the causation of a particular fact or event: "the aggregate of phenomena *existing at any*

moment." And his remarks about an assemblage of conditions encourage us to spot important evidence for the uniqueness in causation. But by the next paragraph, Mill has reverted to thinking in terms of "universals," neglecting what might be discovered to obtain for a single, particular thing: "the *same* phenomena *always* produced by the *same* cause." Being relevantly particular, what may we gain from Mill at his best?

2. CAUSATION BY AN ASSEMBLAGE AND CAUSAL OVER-DETERMINATION

We are clear enough, I take it, about the *objects* of "cause" in sentences relating directly to our principle. But it seems that we might yet advance clarity as regards the terms in the *subject's* position, that is, those terms purporting to denote whatever or whoever caused. In speaking of that which causes, our principle both speaks of some *entity* and also of some *entities*. The singular alternative is appropriate enough. But, in the statement of an alleged *uniqueness* condition, what is the latter, *plural* term doing?

Our principle must not rule out plurality. If it did, it would thereby wrongly rule out such consistent sentences as:

The stone's striking the window and the window's being made of brittle glass caused the window to break.

We want to see, then, how there may be uniqueness in plurality. To do so, we take our cue from Mill, who has already instructed us in such matters: "the difficulty being only that of singling out this one cause from the multitude of antecedent circumstances. . . . The cause indeed may not be simple; it may consist of an assemblage of conditions from which the given effect could result." In this light, let us consider the sentence, displayed just above, concerning the breaking of the window.

As we have remarked, our principle does not rule this sentence inconsistent. Indeed, the principle instructs us as to its consistent interpretation. The subject of the sentence denotes the unique cause of the window's breaking. What caused the window to break is *two* events, in Mill's terms, an *assemblage* of two events. As these events, or this assemblage of them, caused the breaking, nothing else did. In particular, *neither one of these events caused the window to break*. Accordingly, the stone's striking the

window did *not* cause the window to break and, likewise, the window's being made of brittle glass did not cause it. Now, in some *other* case, we may say that a certain stone's striking a certain window caused a certain window to break. But then it follows that, in *that* case, the window's breaking was not caused by the stone's striking it *and* something else, say, the window's being made of brittle glass. It should be clear enough, now, how plurality and uniqueness may be relevantly compatible.

While whatever causes a particular thing may thus be a plurality, we have spotted a logical limit on what may cause it. There is no similar limit, however, on what a particular thing may cause; there is an asymmetry here. Any number of particular things may be caused by a single agent, event or fact; or by an assemblage of these. If an event, for example causes two particular things, then it causes each of them. If a stone's striking a window causes the window to break and John to hear a loud noise, then it causes the window to break, and it also causes John to hear the noise.

I think that things proceed in a relevantly similar manner for other verbs which have causal implications. For verbs that do not, like "precede," for example, matters proceed quite differently. We may best appreciate the difference by employing a Principle of Emphasis. Elsewhere, I have argued for this principle at considerable length.⁷ Briefly, the idea is that by emphasizing key expressions, we do not alter what they mean or express, but we do help ourselves spot logical relations between the concepts they express. To spot the relevant difference between causation and precedence, then, we consider these following two sentences, where in each case our principle has been employed:

The stone's striking the window *caused* the window's breaking *and* the window's being made of brittle glass *caused* it.

The stone's striking the window *preceded* the window's breaking *and* the window's being made of brittle glass *preceded* it.

As emphasis properly brings forth, there is a strong appearance of contradiction with the first sentence, but none at all with the second. There is a uniqueness in causation, but no such uniqueness in precedence.

We may make our points another way, I suggest, by seeing how the word "also" works differently

⁷ Peter Unger, *Ignorance* (Oxford, 1975), Chapter II, Sections 7, 8, and 9.

with "precede" and "caused." While the first of these allows an "also," the second results in contradiction or worse if the word is added:

The stone's striking it *and also* the window's being made of brittle glass *preceded* the window's breaking.

The stone's striking it *and also* the window's being made of brittle glass *caused* the window's breaking.

The reason for this, no doubt, is the relation of these sentences to:

The stone's striking it *preceded* the window's breaking, *and so did* the window's being made of brittle glass.

The stone's striking it *caused* the window's breaking, *and so did* the window's being made of brittle glass.

What is wanted to make the connection explicit, I take it, is an account of the element "so." In any case, as "precede" is wholly acausal, it manifests the symmetry which "cause" does not: Just as a given event may precede each one of any number of others, so that given event may be preceded by each one of any number of other events.

Linguistic intuition thus appears to be on the side of our principle. Acquaintance with the familiar philosophical topic of "causal over-determination," however, may lead many to suppose that our principle must be unsound. They would suppose that cases of causal over-determination don't conform to our principle. But when these cases of causation are properly described, there is no conflict, but only conformity to our principle of uniqueness. A couple of examples will indicate clearly enough how matters run here generally.

Regarding causal over-determination, the simplest sort of case is that where one event causes another, but should the causing event not have occurred, the latter would have occurred anyway, caused by some third event. For example, a certain stone's striking a certain window caused the window to break. But if John had not then thrown that stone, Bill, who was watching things, would have thrown a similar stone, a little bit faster, and the window would have thus been caused to break at the very same time. There is no problem here; for we may say that if the first stone's striking the window did not occur, and so

did not cause its breaking, the second stone's striking the window *would have* occurred, and it *would have caused* the window's breaking. I am not suggesting that thoughts of this case would prompt many to question our condition, but it is good for us to have the case in hand before we come to a case which might.

A more difficult case, it might be supposed, is that where both stones were thrown simultaneously, both striking the window at precisely the same time. But, in fact, there is no difficulty here either. One thing we may say now is that the complex event of the two stones' striking the window caused the breaking. Or, we may say that the breaking was caused by (an assemblage of) *two simpler events*, the first stone's striking it *and* the second's. But, of course, this is not to say that one of these simpler events, or the other, caused anything. Whichever we choose, providing the situation is at all typical, we may go on to say this: If the first stone had not been thrown, the second stone's striking the window *would have caused* it's breaking, while if the second had not been thrown, the first's striking it would have. On the natural assumptions, these are the natural descriptions; on more complex assumptions, we would have other descriptions which conform to our uniqueness condition. Accordingly, whatever problems they may present, cases of causal-overdetermination, however complex, present no problems for our uniqueness condition.⁸

Regarding a plurality of causes, there remains to discuss a line of objection which is a bit more sophisticated than those worries so far encountered. A sophisticated philosopher may easily suppose that our condition necessitates an implausible divergence between the syntax of sentences with the verb "cause" and the semantics of these sentences. For it may be supposed that our condition implies an equivalence to exist between these two sentences:

John caused the window to break.

John caused the window to break by himself.

If there is an equivalence here, the expression "by himself" will make no semantic contribution to the sentence in which it occurs. It is implausible to suppose that such a syntactic element should make no difference semantically; so much the worse for our principle.

⁸ Some recent discussions of these topics include J. L. Mackie, *The Cement of the Universe* (Oxford, 1974), pp. 43-47 and Louis E. Loeb, "Causal Theories and Causal Overdetermination," *The Journal of Philosophy*, vol. 71 (1974), pp. 526-531.

But our principle allows "by himself" to make its contribution. To see the difference it makes, we compare the above two sentences with:

John caused the window to break with someone's help.

Our original unqualified sentence is quite compatible with this last, but the sentence with "by himself" is not. If this third sentence holds, then John caused the window to break, not by himself of course, but with someone's help. Perhaps without the help, he would have caused it anyway; perhaps not; neither is implied by saying that he caused it with the help.

While it will now be granted that we have allowed for the contribution of "by himself," it may be thought that we face a problem on the other side. We cannot, it might be supposed, allow for a contribution of "together" to the semantics of:

John and someone else together caused the window to break.

We need to distinguish this sentence semantically from the unqualified:

John and someone else caused the window to break.

But, given our condition, how are we to do so?

The answer, I think, is to say that the opposite of "together" here will be gotten with "independently," or with an expression to similar effect, rather than with no qualification at all. Accordingly, both of the two sentences displayed just above differ logically from:

John and someone else independently caused the window to break.

The sentence with "together" is inconsistent with the last; the sentence with no qualification is consistent with it and, indeed, is entailed by each of these mutually exclusive qualified sentences.

This is a promising beginning, but what is the function of "together" and "independently" here, so that they solve our problem of semantic contribution. What are they really qualifying? I am far from sure of our answer here, but perhaps I may suggest something. When it's together, then the two people *did something together*—smashed the window with a log perhaps—and *by doing this thing* they caused the window to break. When it's independently, then *each did something independently* of the other—perhaps John threw a stone while the

other person threw another stone. There was no plan for them both to throw. Thus, by each doing his own thing, they caused the window to break.

Such suggestions work as well for various "objects" as for personal agents. For we speak of these objects as doing things. Thus, if the water caused Betty to catch a cold, it might do it by itself or, alternatively, it might do it with help from the cold air. Again, these two might cause John to catch a cold. They might cause it together or, alternatively, independently. In the first case, they might together lower his resistance, thereby causing him to catch the cold. In the second, the water may give him germs while the cold air, independently, prevents him from sleeping.

Where causes are facts or events, our suggestions may appear to find problems. But, then, what do facts do, and what events? Facts obtain; events occur; sometimes in conducive circumstances and sometimes not. If one event causes another *all by itself*, then perhaps it not only causes the latter but does so *without the aid* of conducive circumstances. What about one event and another *together* causing a third to occur? Perhaps it is *by occurring together* that the first two cause the third. In contrast, they might cause it independently. If so, then, perhaps, what happens is this: They each occur independently—at different times, in different places, from different causes, or whatever. And by each thus doing something different, the two may independently cause the third event.

3. TOWARD A SYSTEMATIC ACCOUNT OF "CAUSE"

The concept of causation is no philosopher's invention. To get a handle on it, we need to be sensitive to our language, which embodies it, especially to the word "cause," in terms of which the word "causation" can be defined or understood. We speak of the *causation* of this by that when that is the *cause* of this. Not every sense of "cause," however, is closely relevant to such an understanding, as the following contrast makes clear:

John was the *cause* of the charity's being successful.

The charity was the *cause* which John favored.

In both sentences, the noun "cause" is employed in a very common meaning; but only the meaning in the first sentence is closely relevant to understanding causation. We want an account of the noun in its first meaning which will show how it is related

to the verb "cause." I think that by adhering to our uniqueness condition we get the best, as well as the simplest, account. This will further support the principle I am advancing.

I propose that the basis of a good understanding of "cause," both verb and noun, is found in sentences concerning the causation of particular things:

His smoking caused him to develop cancer.

His smoking was the cause of his developing cancer.

His smoking was a cause of his developing cancer.

On this basis, we may perhaps understand such a sentence as:

Smoking causes cancer,

though exactly what form that extended understanding will take, I am unable to articulate. I do not think this inability need prevent us from thinking we have the proper basis however, nor need it detain us from trying to relate these three basic sentences.

The simplest accounting, I suggest, and the one furthered by our uniqueness condition, relies on identifying *the cause of* a particular thing with *that which caused*, or *is causing*, or *will cause* it. Thus, we may treat the first two of our sentences as logically equivalent. The uniqueness implied by the definite article matches that which our condition picks out for the verb. That this simplest accounting is correct is evidenced by the apparent inconsistency of each of these following four sentences:

His smoking *caused* him to develop cancer, *and so did* something else.

His smoking *was the cause of* his developing cancer, *and so was* something else.

His smoking *caused* him to develop cancer, but (and, so) it *was not the cause of* it.

His smoking *was the cause of* his developing cancer, but (and, so) it *did not cause* it.

We need only now explain the relation of our third sentence, where the *indefinite* article is employed, to the first two. What is *a* cause of an event; what are its *causes*? For *most* cases, I propose that a cause is *a part of the cause*, that it is a part of that which causes the event. Accordingly, the *causes* of an event are the *parts of the cause* of it. In such a case, the cause will be some sort of complex or assemblage. The reason that I qualify my

remarks, speaking only of *most* cases, is that the following appears to me consistent:

His smoking was *a cause* of his developing cancer, *whether or not* it was *the cause*.

But if a cause is always a part of the cause, then it could *never* be the cause itself. Accordingly, we must explain "a cause" in terms a bit more complex. But they must be only a bit more complex. For we may accommodate the troubling case, and adequately paraphrase our third sentence, in terms of a simple disjunction of the two sorts of case:

Either his smoking was the cause of his developing cancer or it was a part of the cause of it.

Other sentences follow suit. Accordingly, if *the causes of* that accident were drunken driving and the slippery road, then drunken driving was *a part of the cause*, and the slippery road was the only other part of it. If the window's being made of brittle glass was *an important cause* of its breaking, then its being made of such glass was *an important part of the cause*.

This account of causes concerns persons as well as events, and indeed anything which may be said to cause a particular thing. Thus, if John was a cause of the charity's being successful, then either he was the cause, in which case he caused it, or else he was a part of the cause, in which case he was a part of that which caused it. If the latter, then perhaps another part of the cause was the economy's booming. If these were the only parts of the cause, then the causes of the charity's being successful were the economy's booming and John. Neither of them caused it, but an aggregate, this time consisting of a person and an event, did cause the charity to be successful. We have hardly presented a complete account of "cause," even limiting ourselves to the relevant sense of the word. But we have, I suggest, made a good beginning in that direction. This further supports our idea of the uniqueness in causation.

Perhaps, this is a good place to remark briefly the relation of our condition to attempts to define causation in terms of (casually) sufficient and/or necessary conditions. While it has recently come under attack, many philosophers have favored the idea that the cause of something must be "sufficient" for its occurrence. It must be something which, at least in the prevailing circumstances, ensures the occurrence of the thing which is thus caused. Along these lines, there has been a tendency for the cause to become "larger and larger,"

with Mill it even swallows up any surrounding circumstances. Such a tendency serves to rule out there being anything else around which might play a part *in causing* the thing. This makes sense on our condition of *one* thing causing, of this being *the* cause, the *whole* cause, which therefore must leave out no part—no contemporaneous necessary condition.

Our uniqueness condition has no need of these conditions of which philosophers speak when seeking to analyze causation. Indeed, I find their words to lead to no conception which is either natural or clear to me. I wish only to help explain their tendency to speak in this particular technical manner. Insofar as our uniqueness condition is of help here, it is further confirmed by the behavior of these philosophers.

4. THE IRRELEVANCE OF ELLIPSIS

Our principle concerns *anything* which might be supposed to cause a particular thing, not only facts, events and people, but physical objects, "universals," and so on. Our principle gives us a strict interpretation for many things we are wont casually to say, and bids us consider that such an interpretation is the correct one. For example, we are wont to assert indifferently:

The slippery road caused the car to crash.

The road's being slippery caused the car to crash.

In fact, however, the road's being slippery is a condition, fact or event, whereas the slippery road is a "physical object." The latter is a road which is, in fact, a slippery one. The road's being slippery is not any road, no matter how slippery that road may be. Accordingly, our principle bids us that at most one of the above assertions is true. I suggest that this is indeed correct, as is evidenced by the apparent inconsistency of:

The slippery road *caused* the car's crashing *and* the road's being slippery *caused* it.

This contrasts with another sentence:

The slippery road *and* the road's being slippery *caused* the car's crashing.

This latter sounds *odd*, I admit; but the point is that it does *not* sound *inconsistent*. Though it does

not concern our principle, perhaps I should say that it seems more likely to me that the road's being slippery caused the car to crash, and indeed quite unlikely that the slippery road itself was the cause.

Let us consider another situation, one which prompts us to say rather indifferently:

Those people caused John to feel depressed.

Indifference caused John to feel depressed.

Those people's being indifferent caused John to feel depressed.

Here, John's feeling is said to be caused by each of three different things: those people, who are certain particular entities; indifference, which is a "universal"; those people's being indifferent, which is a single, particular fact or event. Now, for all our principle says, indifference itself, the "universal," may be the cause, though I very much doubt it. More likely, I think, is that those people are the cause. To my lights, it is more likely still that John's feeling is caused by a certain event, perhaps those people's being indifferent. The point here, of course, does not really concern these relative likelihoods. It is just this: if any one of these is the cause, then none of the others can be.

(A way to allow various things each to cause of course remains open. We may then say, so far as our principle goes, at least, that those people caused John's depression, while their being indifferent caused John to feel depressed.⁹ But I think this purely logical gambit presents little that is likely. More likely in the case is this, I think: Their being indifferent caused John to feel depressed, and it also caused John's depression, which he felt.)

In light of what has gone before, these remarks may seem so obvious as to be not worth making. But serious thinkers, in discussing causation and related issues, have said words to the opposite effect: They have spoken of expressions for the cause as being only elliptical, when these expressions were clearly ill-suited for picking out events or anything much like an event. For one, C. D. Broad writes:

Of course I am well aware that we constantly use phrases, describing causal transactions, in which a continuant is named as the cause and the event in that continuant is mentioned. Thus we say: "The stone broke the window," "The cat killed the mouse" and

⁹ I am here indebted to John Taurek.

so on. But it is quite evident that all such phrases are elliptical.¹⁰

For another, Donald Davidson writes:

... If I poison someone's morning grapefruit with the intention of killing him, and I succeed, then I caused his death by putting poison in his food. . . .

The notion of cause appealed to here is ordinary event-causality, the relation, whatever it is, that holds between two events when one is cause of the other. For although we say the agent caused the death of the victim, that is, he killed him, this is an elliptical way of saying that some act of the agent—something he did, such as put poison in the grapefruit—caused the death of the victim.¹¹

It is clear from his context that Davidson (unfortunately, I think) identifies what someone did with some event. In any event, he is claiming that an assertion like:

John *caused* Jim's death

is elliptical for something like:

John's doing something *caused* Jim's death.

But there is no reason to suppose the long sentence to be elliptical for the short. Is the description of the event elliptical for the simple name "John"? I doubt it.

What this talk of ellipsis does, however, is get us to examine such odd sounding sentences as:

John *and* John's putting poison in Jim's food *caused* Jim's death,

as well as our aforementioned:

The slippery road *and* the road's being slippery *caused* the car's crashing.

Broad and Davidson would, I suppose, explain the admitted oddness in terms of redundancy, or something of the like. But I think that these sentences sound odd because the interpretations needed for their truth concern situations that are, if not downright impossible, which I suspect, at least most bizarre and unusual. I proceed, then, to articulate the interpretations for these sentences.

In the first odd sentence, a certain person, John, is part of the cause. Presumably, it is by doing something that he is. Possibly, it is by putting poison in Jim's food; more likely, it is by doing something else, say, by emptying Jim's antidote

belt of all remedies. This is one way John figures in the cause; *he* is part of it. Another way he figures is this: Perhaps by putting poison in Jim's food, John is involved in a certain event, the event of John's putting poison in Jim's food. This *event*, and *not* John, is the *other* part of the cause of Jim's death.

In the second sentence, the slippery road itself was part of the cause of the car's crashing. Perhaps it was by *doing* something that it was. One thing the road did was curve around very sharply; perhaps it was by doing this that the road was a cause. Or, perhaps, it was not by doing anything, but by simply *being* a certain way, by being sharply curved at a certain point. Now, the road also figures in the cause in another, more indirect way. The road is involved in a certain event or fact, that of the road's being slippery, which is here asserted to be the only other cause of the car's crashing.

When one is thinking of an "object" or a person as causing, or as part of the cause, one will not be likely there and then to be thinking similarly of events involving those very entities. And, the converse is even more obviously the case. Thus, these interpretations, and the sentences interpreted, are clearly no part of everyday thinking or speech. This, I think, is all we need to note in order to explain the oddness of the noted sentences.

Of course, at some deeper level, these sentences, through their interpretations, may prove inconsistent. But to discover this will take argument; it lies beyond the evidence of the ear. However the arguments may run, they will not serve to revive any talk of ellipsis. What our causal sentences are asserting is clear enough for that line to be abandoned. If the sentences are in trouble, that cannot be helped by appealing to ellipsis.

Ellipsis is no more often in place with other causal verbs than it is with "cause" itself. We have already looked at the sentence:

The key opened the door.

This clearly is not elliptical for:

Someone opened the door with the key.

For, as we have seen, it may naturally be used on occasions where we say:

The monkey opened the door.

¹⁰ C. D. Broad, "Determinism, Indeterminism and Libertarianism" in his *Ethics and the History of Philosophy* (London, 1952); reprinted in Bernard Berofsky (ed.), *Free Will and Determinism* (New York, 1966), where the quoted passage is on p. 157.

¹¹ Donald Davidson, "Agency" in *Agent, Action and Reason*, ed. by Robert Binkley, Richard Brunaugh, and Ausonio Marras (Toronto, 1971), p. 10.

We might then think that it is elliptical for:

Someone or something opened the door with the key.

But where is the plausibility in this?

Let us suppose that after the monkey used this key to open the door, the monkey then pulled the key out of the door and let go of it. The key fell to ground, we may then suppose, at least as surely as we might suppose it true that the key really opened the door. Now, if something did *X* and it also did *Y*, then, I suppose, it is true to say that it did *X* and *Y*. Can we truly say here that the key opened the door and fell to the ground? I think not. If the monkey, after doing the rest, then fell to the ground, we could truly say that the monkey opened the door and fell to the ground. Why are we reluctant in the case of the key? Whatever more is going on, surely part of the reason is this: We don't really believe that the key opened the door.

5. THE INTRANSITIVITY OF THE CAUSAL RELATION

Our principle governs the causation of facts or events, not only by other particular facts or events, but also by persons, objects, universals, assemblages of these, and whatever. But the case of one fact or event causing another has been of particular interest to many philosophers. This is largely because this case rather naturally submits to simple questions framed in the language of "classical logic." In this vein, our condition provides an unusual answer: Given the irreflexivity of this causation, which is all but impossible to deny, the causal relation is actually *intransitive*. What does this mean?

When it is asserted that an event *A* causes an event *B*, we may take it as being asserted that a dyadic (causal) relation holds between *A* and *B*, with *A* in the first place, that of the cause, and *B* in the second, that of that which is caused. This way of thinking is fully general. Accordingly, considerations of truth aside, it may be asserted that a given event, *A*, bears the relation to itself, that is to assert that *A* causes itself. In classical logic, there are three familiar "dimensions" for classifying dyadic relations: Those of reflexivity, symmetry and transitivity. In terms of these dimensions, what may we say about our causal relation?

It is quite clear, I suggest, that no event can cause itself; the causal relation is neither reflexive nor even non-reflexive; it is *irreflexive*. The stone's striking the window can't cause the stone to strike the window, nor can the window's breaking cause the window's breaking. A stone's striking a window on one occasion might cause it to strike that window on another, presumably later, occasion. But this is not an event causing itself; rather, it is an event which answers a certain description causing another which answers that same description. A finer description, implying uniqueness, will prevent such a confusion from arising.

The asymmetry of any dyadic relation will entail its irreflexivity, though the converse doesn't hold. This is entirely familiar. May we assert only the irreflexivity of the causal relation? Or, may we assert its asymmetry, from which that irreflexivity may then be presumed to follow? It is quite clear, I believe, that the causal relation is even *asymmetric*. If an event *A* causes an event *B*, then *B* cannot cause *A*. Of course, *B* might cause an event entirely similar to *A*, but then that event will be distinct from the event *A* itself. If the stone's striking the window caused the window to break, then the window's breaking did not cause the stone then to strike it.

It should be observed that this point about asymmetry is not any point about "backwards causation." Whether a later event can cause an earlier one is a different matter. Asymmetry alone cannot rule out the later breaking of the window from causing the earlier striking by the stone. Rather, asymmetry will say that if the breaking *does* cause the earlier striking, then the striking doesn't cause the breaking. Whatever may be the situation concerning causation and time, I can't see how one can properly deny the asymmetry of the causal relation.

We may also divorce from thoughts about time the question of the dimension of transitivity. But, in any event, here matters are more difficult. Some philosophers hold that the causal relation is transitive.¹² For them, if our stone's striking our window caused the latter to break, and the window's breaking caused John to feel annoyed, the stone's striking the window caused John to feel annoyed, although perhaps indirectly. Other philosophers hold that such causing is not transitive, but not

¹² A recent example of a philosopher who favors transitivity is David Lewis in his "Causation," *The Journal of Philosophy*, vol. 70 (1973), p. 563. (Unusually, on p. 565, Lewis seems to hold that causation is not reflexive, but is non-reflexive).

intransitive either, and so that it is a non-transitive relation.¹³ For them, our stone's striking the window *may* cause John to feel annoyed and, again, it *may not*; which it does will, presumably, depend upon various circumstances of the particular case in question. I am familiar with no prominent philosopher who has held that the causal relation is intransitive, that in our situation, given the first two conditions, it *follows* that the stone's striking the window did *not* cause John to feel annoyed. But only such a philosopher would be right about the causal relation.

With our uniqueness condition in hand, the derivation of this unpopular consequence is neither long nor complex. We are to consider a situation where an event *A* causes an event *B*, and where *B* causes an event *C*. Might *A* cause *C*? As *B* causes *C*, nothing other than *B* itself will cause *C*. That is what our principle says. Thus, *A* will cause *C* *only if* *A* is identical to *B*, that is, only if *A* is really the very same event that *B* is. Might *A* be the same as *B*? We have said that *A* causes *B*. As causing is irreflexive, *A* does *not* cause *A*. Thus, *A* and *B* are distinct, and not the same. As the two are different, while *B* causes *C*, *A* does not. Therefore, the causal relation is *intransitive*.

Our derivation is straightforward enough. What is the truth? We consider test sentences:

The stone's striking the window *caused* (the onset of) John's feeling of annoyance *and also* the window's breaking *caused* it.

The stone's striking the window *preceded* (the onset of) John's feeling of annoyance *and also* the window's breaking *preceded* it.

If one makes no move to "explain away" the data, one will agree that, while there may be transitivity in the relation of precedence, there is only intransitivity in the relation of causation.

Indeed, in this spirit we may do well to consider a problem posed by David Lewis.¹⁴ What are we to make of such sentences as "Only driving errors cause auto accidents"? If we deny intransitivity, as Lewis does, we will have a problem. What about cases of excessive drinking which cause driving errors—don't they *ever* cause auto accidents? Transitivity, and presumably most mere non-transitivity, will feel constrained to say "Yes." But, then, they will have to interpret the original "only" sentence in some such labored way as this: "Only causal histories that involve driving errors are ones that, in the place of that which is caused, contain auto accidents." Can this be what is said by the simple and apparently innocent: "Only driving errors cause auto accidents"? It would seem not. And our account says as much. For, on our account, those cases of excessive drinking, while they cause driving errors, *never do* cause auto accidents. Only driving errors *cause* auto accidents. Of course, we want a detailed account of this sentence, which we do not as yet have. But it seems that only intransitivity, and our uniqueness condition, will allow for a detailed account which is most plausible.

New York University

Received March 3, 1976

¹³ A recent example of a philosopher who favors non-transitivity is Michael Scriven in his "Causation as Explanation," *Nous*, vol. 9 (1975), p. 13. (Unlike Lewis, Scriven here holds causation to be irreflexive. But, almost equally unusual, he does hold causation to be non-symmetric.)

¹⁴ Lewis, *op. cit.*, p. 558. I am grateful to Lewis for pointing out to me how my account most simply handles this problem.

II. PSYCHOPATHY AND MORAL UNDERSTANDING

ANTONY DUFF

I
CRITICS of contemporary psychiatry argue that concepts of mental disorder are wrongly used to exclude from ordinary human rights and relationships people who are fully rational and responsible; that the pseudo-medical terminology of disorder gives a spurious air of objectivity to our *value*-judgments on those whose conduct we find unacceptable, and disguises the fact that we are *imposing* on them our own moral and social values.¹ Criticism has been fiercest of the concept of "psychopathy": even critics who allow the concept of mental disorder to be legitimately extended beyond the narrow limits of the M'Naghten Rules dismiss "psychopathy" as a pseudo-medical moral judgment.² They argue that this "diagnosis," based solely on the anti-social conduct it is meant to explain, is vacuously circular; that it marks only our disapproval of the conduct, not the medical discovery of its origin in a disordered condition of mind; that we cannot distinguish a disordered psychopath from a rational criminal or rebel.

My aim is not to defend current usage of the term "psychopathy"—it is certainly too often used to ascribe disorder to those whom we should rather call immoral or unusual—but to show that it *can* be used to identify a kind of disorder not captured by the more readily accepted criteria, and thus that critics who argue, not just that it is abused in practice, but that it has *no* legitimate use, are wrong. The argument is also relevant to wider philosophical and moral issues. As part of a more general attempt to understand the problematic concepts of mental health and disorder, it should help to provide a sounder theoretical basis for our treatment of the disordered. It is relevant too to philosophical dis-

cussions of responsibility and rationality: an account of mental disorder, as a condition in which a person's rational capacities, and thus his status as a rational agent, are impaired, requires a correlative account of what it is to *be* a rational and responsible agent. More directly, a discussion of psychopathy will demonstrate the inadequacy of one traditional philosophical account of practical rationality, which underlies many criticisms of the concept of psychopathy. I hope to show at least that these ambitious claims are well-founded.

An account of psychopathy as a disorder should begin with a general account of the concept of mental disorder. I can here only assert what should be argued: that a person is mentally disordered insofar as he suffers some abnormal and harmful impairment of his rational capacities.³ The identification of impaired capacities provides the "objective" basis of judgments of disorder: the requirement of harmfulness indicates the evaluative aspect of such judgments.⁴ We must thus identify the capacities essential for a rational human life, and provide criteria for their assessment. Two such capacities can be readily identified: the intellectual capacity for reasoning about the empirical features of one's environment, one's actions, and their consequences; and the capacity to control one's actions, and resist contrary impulses, in the light of one's rational purposes. We can thus identify two kinds of disorder, involving the impairment of these capacities. Serious intellectual incapacity, typical of severe mental deficiency and delusional psychoses, brings the agent under the strictest interpretation of the M'Naghten Rules: he is "labouring under such a defect of reason . . . as not to know the nature and quality of the act he was doing."⁵ An impaired

¹ C.f. Barbara Wootton, *Social Science and Social Pathology* (London, 1959), chs. VII–VIII; Thomas Szasz, *The Myth of Mental Illness* (New York, 1961); Antony Flew, *Crime or Disease?* (London, 1973).

² Besides those noted above, c.f. Anthony Kenny, "Mental Health in Plato's Republic," *Proceedings of the British Academy*, vol. 55 (1969), pp. 249–253; Nigel Walker and Sarah McCabe, *Crime and Insanity in England* (Edinburgh, 1973), ch. 10.

³ For accounts of disorder based on the notion of incapacity, c.f. Antony Flew, *op. cit.*; Jonathan Glover, *Responsibility* (London, 1970), ch. 7.

⁴ I would argue that we can give a parallel, and equally evaluative, account of the concept of physical disorder, in terms of the harmful impairment of capacities: c.f. Antony Flew, *op. cit.* pp. 35–53; Jonathan Glover, *op. cit.*, ch. 6.

⁵ The Rules are quoted in full in the *Report of the Royal Commission on Capital Punishment 1949–53* (London, 1953), Appendix 8: this strictly intellectualist interpretation is the one which has been officially adopted by English courts; c.f. *R. v. Codere*, (1916) 12 C. A. R. 21; *R. v. Windle*, (1952) 2 Q.B. 826.

capacity for control subjects the agent to "irresistible impulses"—to compulsive actions which he rationally rejects but cannot prevent himself doing.⁶ The abnormal impairment of these essential human capacities constitutes a disorder.

To focus on our central problem, I will define a psychopath, negatively, as one in whom *these* capacities are unimpaired: he can discover the empirical features of his environment and actions, and he acts as he wants to. I will argue later that we can legitimately extend the notions of rational understanding and control, and say that a psychopath is unable to understand the nature and quality of his acts, or to control them rationally: but this extension must be justified. I will try to construct an account of psychopathy as a disorder, partly by showing the inadequacies of some suggested accounts, and partly by showing how the inadequacies of traditional empiricist accounts of the notions of understanding and rationality in moral and emotional contexts may hinder the recognition of this kind of disorder.

II

Accounts of psychopathy usually begin, and sometimes end, with accounts of psychopathic behaviour.

"The psychopath . . . is constantly getting into trouble with neighbours, 'friends', family, and police. Periods of socially acceptable conduct alternate erratically with destructive pranks, petty thefts, forgeries, fights, intoxication, miscellaneous other minor and often pointless crimes. Inexplicable minor cruelties, humiliations, inconveniences, are dealt out freely by him. He knows the probable consequences of his acts . . . but does not care—very often does not care even enough to try to protect his liberty by concealing his doings."⁷

But such accounts of criminal and anti-social behaviour, injurious to others and apparently to himself, provide by themselves no proof of disorder. Persistent criminality, for instance, could be a criterion of disorder only within a determinist perspective which, denying responsibility to anyone, takes the fact of crime as proof that "something has gone wrong." But the sense of "disorder" is now radically changed: for we cannot assume that all

criminals are less capable of rational understanding and control than the rest of us. Furthermore, if crime is *the* criterion of psychopathy, the label becomes utterly uninformative: it neither explains nor illuminatingly redescribes criminal conduct, but simply marks its occurrence; it precludes any distinction between psychopathic and other criminals.⁸

Nor will it help to provide a fuller account of "psychopathic" patterns of behaviour, specifying the kinds of offence committed—violent, sexual, etc; adding that his conduct is generally anti-social, not just criminal; that he exhibits no concern either for his own interests or for those of others, acting without any recognisably rational motive, making no attempt to avoid detection, and destroying the ends which prudence is thought to dictate, and the relationships which men value; that, in short, his conduct is by our standards grossly immoral, imprudent, and irrational. This distinguishes a psychopath from other criminals: and *if* we can suppose that he shares our conceptions of prudence and value, he must be disordered; conduct so obviously destructive of the agent's own ends must be attributed to an inability either to realise its obvious consequences, or to resist impulses which *he* sees as irrational. But why suppose this? He clearly *knows* what he is doing (he can tell us), and is *not* trying to resist his impulses: he just has values very different from ours.⁹ Human values, both moral and prudential, are irreducibly diverse: a man can rationally live by values very different from, and perhaps repugnant to, our own. An adequate concept of disorder must allow a distinction between the rebel against conventional values and a disordered psychopath: mere difference, in conduct or values, precludes this distinction.

But could the psychopath help acquiring his perverse values; could he now change them? This test, if intelligible, will not distinguish a psychopath from other criminals or rebels.¹⁰ If it means "could he acquire other values *if* he wanted to?", it asks whether he can modify his derivative purposes in the light of his more basic desires: could he lead a law-abiding and conventional life if he thought it better to do so? But the psychopath, *if* he wanted to,

⁶ There are, however, more problems with this category of disorder than is sometimes recognised: c.f. Antony Flew, *op. cit.*, pp. 59–60; Jonathan Glover, *op. cit.*, pp. 97–101 and 135–140, for discussions that are not really adequate.

⁷ Herbert Fingarette, *On Responsibility* (New York, 1967), pp. 25–26. C.f. also Hervey Cleckley, *The Mask of Sanity* (St. Louis, 1964); D. K. Henderson, *Psychopathic States* (London, 1939); William and Joan McCord, *The Psychopath* (New York, 1964).

⁸ C.f. Antony Flew, *op. cit.*, pp. 73–78; Barbara Wootton, *op. cit.*, pp. 249–250.

⁹ C.f. Vinit Haksar, "The Responsibility of Psychopaths," *The Philosophical Quarterly*, vol. 15 (1965), pp. 140–143.

¹⁰ C.f. Vinit Haksar, "Aristotle and the Punishment of Psychopaths," *Philosophy*, vol. 39 (1964), pp. 81–88.

could change his ways as easily, or as arduously, as anyone else. And it is meaningless to ask whether he could *want* to change: any intelligible notion of capacity must be explained hypothetically, as a capacity to do something *if* I want to; at the level of those basic desires from which all others, including the desire to change, are derived, notions of capacity and incapacity become meaningless. Responsibility cannot require the capacity to change our *basic* desires or values: for we try to change desires, and thus reveal a capacity or incapacity to do so, only on the basis of more basic desires, which are the unchosen *given* of all practical reason and rational action.¹¹

Two other suggestions may be mentioned. Firstly, that the fact that the psychopath is not deterred from crime by punishment, or from anti-social and irresponsible conduct by rational criticism, provides good Utilitarian grounds for holding him non-responsible: for a man is responsible insofar as we can usefully modify his conduct by rational deterrence and criticism; if these persistently fail, we must seek other ways of controlling his conduct, such as therapy or detention, and count him disordered—as insusceptible to rational persuasion.¹² This suggestion may serve the Utilitarian goals of deterrence and prevention, so long as the alternative to prison is enforced psychiatric treatment: but it does not provide an adequate criterion of disorder. We must distinguish the man who is undeterred or unpersuaded because he is rationally unconvinced by our arguments, or is committed to some principle which justifies his actions, from the man who is rationally incapacitated, and thus disordered: that we cannot *in fact* deter or persuade a man does not show him to be incapacitated or less than fully rational; persistent crime may still be rational.¹³

Secondly, it has been suggested that the correlations between behaviour independently identified

as psychopathic and the presence of certain physiological abnormalities (an extra Y chromosome; abnormalities in EEG patterns and capillary loop formations) can justify a diagnosis of disorder.¹⁴ But even a 100% correlation between such behaviour and such statistical abnormalities could provide evidence of disorder *only* if the abnormalities could also be correlated with some independently identifiable incapacity.¹⁵

These preliminary attempts to provide criteria by which a psychopath can be judged disordered are inadequate. But some point towards the truth: it is by asking whether a psychopath exhibits any intelligible values in his own life; how he is related to those values of ours which he does not share; how far any rational argument is possible with him (though not just whether we can in fact persuade him by such means); that we will identify his disorder, and the crucial differences between him and the rational rebel or criminal.

I will argue that we cannot find in a psychopath's life any intelligible dimension of value, emotion, or rational concern: and that this lack is logically connected to an inability to understand this dimension of the lives of others; the values, interests, and emotions which inform their lives and actions; the moral, prudential, and emotional aspects of his own actions. I will thus be arguing that there is a close logical connection between two commonly identified features of psychopathy: firstly, that it involves an incapacity for such emotional and moral responses as love, remorse, and concern for others:

"He can commit the most appalling acts yet view them without remorse. The psychopath has a warped capacity for love. His emotional relationships, when they exist, are meagre, fleeting, and designed to satisfy his own desires. These last two traits, guiltlessness and lovelessness, conspicuously mark the psychopath as different from other men."¹⁶

¹¹ C.f. P. H. Nowell-Smith, *Ethics* (Harmondsworth, 1954), ch. 19.

¹² C.f. William Kneale, "The Responsibility of Criminals" in H. B. Acton (ed.), *The Philosophy of Punishment* (London, 1969), pp. 185–190.

¹³ This suggests that a strictly Utilitarian account of responsibility will be inadequate: indeed, I would argue that a strictly Utilitarian approach must lead us to abandon the notion of responsibility altogether, and with it the related distinctions between punishment and therapeutic treatment. C.f. Barbara Wootton, *Crime and the Criminal Law* (London, 1963) for a consistent working out of such an approach.

¹⁴ C.f. D. K. Henderson and R. D. Gillespie, *Textbook of Psychiatry*, 10th edition, revised I. R. Batchelor (London, 1969), p. 531; David Stafford-Clark, *Psychiatry Today*, 2nd edition (Harmondsworth, 1963), p. 117; *Evidence of Dr. Stafford-Clark and Dr. Hill to the Royal Commission on Capital Punishment* (1949–53), *Minutes of Evidence*, vol. XIII, 297ff.

¹⁵ C.f. Antony Flew, *op. cit.*, pp. 76–78. We should note though that these data may at least be *suggestive*, since the EEG patterns and capillary loop formations of psychopaths are very like those of children, which may lead us to look for some kind of immaturity or deficiency in the psychopath.

¹⁶ William and Joan McCord, *Psychopathy and Delinquency* (New York, 1965), p. 14; quoted by M. S. Pritchard, "Responsibility, Understanding, and Psychopathology," *Monist*, vol. 58 (1974), p. 631.

Secondly, that a psychopath, although not intellectually incompetent, is unable properly to understand the "nature and quality" of his acts, since he cannot grasp those emotional and moral aspects which are as much a part of them as their empirical features.¹⁷ A psychopath is not a rebel, who *rejects* more conventional values and emotions in the light of some favoured conception of the good: he is a man who has never come to understand, or to share in, this dimension of human life.

Here, however, we face a familiar philosophical argument, based on an empiricist account of practical reasoning and on the supposed distinctions between fact and value, reason and emotion, and designed to show that we can call a man disordered *only* if he is intellectually incompetent or suffering some strictly irresistible impulse.¹⁸ On this account, the values, wants, and emotions which determine and explain a man's actions are analysable into sets of factual beliefs about the world, his actions, and their consequences, to which are joined certain attitudes or reactions—of approval or disapproval, liking or disliking. Rational assessment, and thus a judgment of disorder (which must rest on the discovery of some serious irrationality or rational incapacity), is possible only of factual beliefs, or of the *internal* consistency of a man's wants and values: we can regard him as rationally defective, or disordered, only if we can identify some irrationality of factual belief, or some radical conflict between his actions and his wants, or between the wants on which he acts and other wants which we can say are more important to him. But no such rational criticism is possible of those basic reactions and attitudes which form his basic wants and values: for they are neither true nor false, nor are they means to any further end. A man is rational and responsible just so long as he knows, factually, what he is doing, and is doing what he wants.

On this account there is nothing wrong with a psychopath: he knows, factually, what he is doing; he does what he wants, and would not do it if he did not want to; he has emotions, values, and wants, since he clearly has pro- and anti-attitudes towards actions and states of affairs. He does not *share* our values and emotions: but since we cannot say that

he wants, but is unable, to share them, we cannot regard him as either irrational or incapacitated on this basis; for these basic values and emotions are neither rational nor irrational—they simply *are*. Nor can we say that he cannot understand our values and emotions: for understanding requires only that he be able to grasp our factual beliefs and observe what our pro- and anti-attitudes are; and this an intellectually competent psychopath, who has pro- and anti-attitudes of his own, can do. If someone claims that "real" understanding of values and emotions requires that we actually share them, then a psychopath cannot understand ours, nor we his: but the same will be true of anyone whose values and emotions we do not share; thus we cannot distinguish a psychopath, as disordered, from a rational rebel or criminal. In any sense which could be relevant to the ascription of responsibility or disorder, a psychopath can understand his actions and our values.

My aim in the rest of this paper will be to show that, although a psychopath satisfies *these* criteria of practical rationality, he is still disordered, because incapable of a proper understanding of those values, interests, and emotions which he so conspicuously does not share: in showing this, I will also be showing the inadequacy of this account of practical reasoning and understanding.

III

A more detailed example will assist this task. Meursault, in Camus' *L'Etranger*, seems a paradigm example of a psychopath.¹⁹ Although the prosecutor portrays him as a callous and cold-blooded criminal, and the defence counsel as a "decent hard-working young man" who "for one tragic moment lost his self-control," it is clear that they both read into his conduct purposes, interests, and values which are not there. He does not help Raymond, or kill the Arab, in pursuit of a calculated vendetta: he helps because he is asked, and can see no good reasons for or against it; the killing is almost something that happens to him, not something he does with any intention or purpose. Nor does he exhibit the kinds of concern appropriate

¹⁷ C.f. Hervey Cleckley, *op. cit.*; compare too the extended interpretation of the M'Naghten Rules favoured by Australian courts: Norval Morris, "The Defences of Insanity in Australia" in Gerhard Mueller (ed.), *Essays in Criminal Science* (New Jersey, 1960), pp. 273–298.

¹⁸ Such an account clearly underlies Flew's argument (Antony Flew, *op. cit.*), and is closely related to the argument about capacities mentioned above, pp. 190–191.

¹⁹ Albert Camus, *L'Etranger*, translated as *The Outsider* (tr. Stuart Gilbert) (Harmondsworth, 1961); c.f. J. G. Murphy, "Moral Death: A Kantian Essay on Psychopathy," *Ethics*, vol. 82 (1972), pp. 288–89.

either to a rational egoist or to a "decent hard-working young man": he has no interest in promotion at work, or in developing lasting emotional relationships of love, or friendship—indeed, he finds these notions meaningless; he makes no effort to avoid conviction or sentence; he feels neither remorse for his crime nor pity for himself. His life lacks that dimension of emotion, value, and interest which gives cohesion and intelligibility to other lives, connecting a man's actions to each other and to him.²⁰ It consists of discrete episodes and impulses, limited to the present moment and the immediate context; the absence of recognisably rational purposes and values makes it difficult sometimes even to say that he *acts*, rather than that events and acts happen to and through him.

However, it may be suggested that Camus shows Meursault's to be an intelligible life, lived by a rational, though unusual, conception of human life and its possibilities: for Meursault's revelation at the end of *L'Etranger*, supplemented by *The Myth of Sisyphus*, provides an intelligible metaphysical and ethical background to such a life. Given his account of human life as a solitary movement towards death; of the impossibility of discovering any meaning within or without the world: given a fundamental ethic of honesty—of living in the light of the truth, however harsh: we can understand someone who regards as illusory the concerns of others for marriage, jobs, and the future; who, abandoning the futile attempt to build lasting relationships, and regarding others as of no more significance than himself, lives strictly for the moment, seeking only to extract the maximum intensity from each passing experience. Meursault has *seen through* the illusory concerns with which others surround and support themselves: it is honesty, not incomprehension, which explains his failure to act as others expect, or to exhibit the kinds of concern—for his mother, for Marie, for his victim, for himself—whose absence we find so puzzling or repugnant. He is a rational rebel: we may condemn or admire him, but cannot call him disordered.²¹

This attempt to show a life like Meursault's to be intelligible can succeed only if it can connect such a life to a shared metaphysical and ethical tradition, showing his beliefs and values to form an in-

telligible extension, development, or modification of concepts and values within a tradition we share with him.²² But, even if Camus succeeds in providing an intelligible background which *could* thus make sense of a life like Meursault's, he still fails to render *Meursault's* life intelligible. For we *cannot* find in Meursault's life the kind of understanding, insight, and concern for honesty which, at the end of the book, we are asked to see in it: he exhibits, not an insight into a truth which others conceal from themselves and a dedication to living in the light of that truth, but rather a blank *incomprehension* of the values, interests, and emotions which inform the lives of those around him, and a complete absence of any conception of truth or value of his own.

He is not intellectually defective: he can give a factual account of his actions; he can see *that* others find such matters as death, love, friendship, career, important, and that they are irritated or annoyed by his actions and responses. But his understanding is still deficient: for he cannot see *how* these things can be important, how they can provide reasons for action and judgment; he cannot understand the emotional and moral significance these aspects of life have for others. And thus he cannot understand the "nature and quality" of his actions, since he has no grasp of these aspects of them. He has not *seen through* the illusions of others, since he lacks the understanding of them which that would require: it is not a concern for honesty—a refusal to lie—which explains his refusal to co-operate with his counsel, since he is incapable of the moral understanding which such an ethical stance requires. He is not a man living by unconventional or unusual values: for he has no rational values, concerns, or interests at all; that dimension of thought and experience which he cannot understand in the lives of others is equally missing from his own life. He is more like a mental defective than a psychotic: his grasp of values and concepts is *lacking*, not distorted.

My claim is that Meursault's life is psychopathic, and must remain so unless we can show, not just that a life *like* that could be rendered intelligible by providing an appropriate background, but that *his* life, as lived by him, has that kind of meaning. To

²⁰ My understanding of *L'Etranger* has been much helped by S. R. Sutherland, "Imagination in Literature and Philosophy," *British Journal of Aesthetics*, vol. 10 (1970), pp. 261–274.

²¹ Compare Mailer's account of the "White Negro's" psychopathic life: Norman Mailer, "The White Negro" *Advertisements for Myself* (London, 1969), pp. 269–289.

²² C.f. R. W. Beardmore, *Moral Reasoning* (London, 1969), ch. 5.

substantiate this claim, I will try first to explicate the kind of understanding of moral values which Meursault lacks, and the connection between such understanding and an acceptance of, or concern for, these values, which will involve expanding my brief comments above on the conditions for the intelligibility of moral beliefs. I will then discuss the way in which this lack of moral understanding is logically connected to his failure to understand the more generally practical and emotional aspects of his actions and of the lives of others, and the logical connections between this lack of understanding of, and concern for, others and the absence in his own life of any conception of, or concern for, values and interests of his own, and the related absence of emotional responses and concerns.

IV

Psychiatrists and philosophers sometimes identify a psychopath's "lack of conscience" as grounds for calling him disordered.²³ Of course, most of us are at various times more or less indifferent to the moral aspects of our actions, and to the interests of others—for which we are properly blamed: but a psychopath is not just *sometimes* thoughtless or indifferent; he *never* exhibits any sincere concern for the interests of others, or remorse for his actions. But why should this render him disordered, rather than just more callous or wicked than others?

Can we say that such *complete* lack of concern for the values we recognise is by itself enough to render him disordered: that our moral reactions to, and relationships with, each other as responsible moral agents are possible only within a common form of life, based on shared values; that we cannot enter into such moral relationships with a man whose values are so alien to ours, and thus cannot properly hold him responsible?²⁴ But to avoid the response that lack of concern is grounds for, not a defence against, condemnation, we must also show how a psychopath's lack of concern involves a lack of understanding, which is connected to the absence in his own life of *any* intelligible values. M. S. Pritchard has argued that such a lack of concern for moral values *amounts* to an inability to understand them.²⁵ He claims that moral communication

and understanding are possible only within a shared form of life, which involves "agreement not only in definitions but also in . . . judgments," reactions, and emotions: and thus that an understanding of moral concepts and values requires not just an intellectual recognition of the criteria by which others make moral judgments, but a *concern* for such values, which enables me to understand these judgments as *practical*.²⁶ His argument points the way to a better understanding of psychopathy as a disorder: but he gives a misleading account both of the kind of concern required for moral understanding and of the kind of "intellectual" understanding possible for a psychopath.

We must allow that we can understand not only values with which we profoundly disagree, but also values belonging within forms of life other than our own, such as those of historical or alien cultures, with which there is no question of either agreement or disagreement: we can gain not only an "intellectual" understanding of such values, but a moral understanding of their meaning, and of how they can intelligibly provide reasons for action. Understanding requires neither a common form of life nor any *identity* of moral concerns with those we wish to understand. It *does* require that we have moral values, and a moral language, of our own; *and* that we can find logical connections between out concepts and values and theirs: we come to understand them by tracing out these connections. Unless we can understand their concepts and values in terms of our own (which may involve extending, stretching, and modifying ours), we cannot understand them at all: unless we have moral concerns of our own, values to which we are practically committed, we cannot understand their values as practical.

Pritchard rightly emphasises that coming to understand values must involve coming to care for *some* values: a child learns moral concepts and the language of values, as an agent and participant; not just as an observer. He does not (and this is a logical, not a psychological, point) learn first to identify a class of actions under some neutral description, and *then* discover that people also adopt a (logically distinct) attitude of approval or disapproval towards such actions: he is taught to

²³ C.f. the works cited above by Cleckley, McCord & McCord, Fingarette, and Pritchard; also G. M. Stephenson, *The Development of Conscience* (London, 1966).

²⁴ C.f. Herbert Fingarette, *op. cit.*

²⁵ Pritchard, *op. cit.*

²⁶ Pritchard draws on Wittgenstein and A. E. Murphy for his account of practical reason: Ludwig Wittgenstein, *Philosophical Investigations* (Oxford, 1958); A. E. Murphy, *Theory of Practical Reason* (LaSalle, 1965).

identify them *as* right or wrong, as actions which he has reason to do or not to do.²⁷ The major defect of empiricist accounts of moral concepts is that they suppose we can distinguish moral attitudes from the kinds of action towards which they are directed; and thus that an understanding of a moral view requires only a grasp of the descriptive criteria through which the attitude is directed towards a class of actions. But Pritchard's insistence on concern as a requirement of understanding ignores the fact that, having learned to understand (in part by learning to care for) certain values, a rational agent becomes able not only to criticise and modify the values he has been taught, but also to understand values very different from them.

Moral understanding is not a purely intellectual matter: but it can be exhibited by someone who remains unmoved by the values he seeks to understand. Indeed, the kind of "intellectual" understanding which Pritchard allows to the man "lacking a sense of justice," is, as he describes it, a moral understanding:

"If he is sufficiently observant and cautious in his behaviour, he might acquire a reputation as an astute judge of issues of justice. He might learn very well what others count as relevant in settling these issues. He could feign guilt feelings or act apologetic when criticised for acting unfairly. He could appear indignant when injustices are done to others by others and resentful when they are done to him. And he could put on a show of other-regard and positive concern to be just. Still, since all of this is only mimicry, he would lack a sense of justice. And, although he would not lack 'information' or 'data' that others have, he would not see what they see."²⁸

It is often said that an intelligent psychopath can purvey an impression of moral understanding and concern; that it is only his unguarded actions and responses which give him away.²⁹ But what Pritchard describes *is* a moral understanding, requiring more than purely intellectual capacities, though less than an actual concern for *those* values: it requires emotional and imaginative capacities which a psychopath lacks. To call it a purely intellectual understanding, requiring only a grasp of the *descriptive* criteria by which the application of moral terms, and the propriety of moral responses, are de-

termined, revives the very distinction between fact and value which this argument should lay to rest.

For what does the competent and astute discussion of moral issues involve? It involves more than the ability to apply learned formulae which provide descriptive criteria for the identification of actions to which moral labels can then be attached—of this an intelligent psychopath may well be capable. A person who is to exhibit moral understanding must also be able to explain and criticise these moral rules—which involves more than showing how they do or do not derive from other formulae; he must be able to show how these rules may or may not be *extended* to cover new cases, which do not fall exactly under any specified set of descriptive criteria; to discuss rationally the resolution, or the impossibility of resolving, cases of conflict. Following moral rules, the ability to "go on in the same way,"³⁰ requires more than the intellectual capacity to acquire and apply fixed formula: it requires a *creative* capacity to understand the significance of the value in question, and to discuss, extend, and criticise its application.

This capacity, or its lack, can be exhibited in detached discussion of values which we do not share: its exhibition, in creative and imaginative discussion, *constitutes* a moral understanding of those values. But it requires a kind of sensitivity and imagination possible only for someone who already shares in some form of moral and emotional life, whose life includes values and emotions logically connected to those he is trying to understand: it is this which a psychopath, like Meursault, lacks. He may learn, for instance, that it is wrong to hurt people, and notice that this word, and this moral label, are applied not only to physical assaults, but also to some ways of talking to others, and to failures to do what we said we would do. He may thus build up a set of descriptive criteria for the application of the term "hurting," and for the judgment that it is wrong. But understanding requires that he also be able to explain the connections between these criteria; and to see how the term could be extended to cover new cases which do not fall under this list—which will count as cases of hurting not (just) because of their empirical or descriptive resemblances to the more

²⁷ C.f. D. Z. Phillips and H. O. Mounce, *Moral Practices* (London, 1969), ch. 1; Richard Norman, *Reasons for Actions* (Oxford, 1971), ch. 3.

²⁸ Pritchard, *op. cit.*, p. 640.

²⁹ C.f. Stephenson, *op. cit.*, pp. 19–20.

³⁰ C.f. Wittgenstein, *op. cit.*, I.185ff; Peter Winch, *The Idea of a Social Science and its Relation to Philosophy* (London, 1970), pp. 24–33.

central cases, but because of the emotional and moral connections we can see between them. And of this a psychopath is incapable: lacking any dimension of emotion or value in his own life, he cannot see how, though he may be told that, it can be hurtful to someone, and a kind of dishonesty, not just to fail to do what you explicitly said you would do, but also to fail to do something you know they have come, tacitly, to rely on you to do; nor could he argue that this is morally different from other kinds of injury.

A psychopath's incapacity, his lack of understanding, is revealed by the fact that he is capable of no more than a stereotyped and rigid application of moral formulae he has learned.³¹ His relation to moral concepts is like that to aesthetic concepts of a person lacking in aesthetic sensitivity (which is distinct from an actual concern for the artistic objects he is discussing): he can parrot certain stereotyped judgments, and apply them to new cases which fall under such descriptive criteria as he has learned; but he can neither understand, nor intelligently discuss, criticise, or extend the rules he has picked up.

I have been trying in this section to steer a way between two untenable extremes: the view that moral understanding requires only an intellectual grasp of the descriptive criteria for the application of moral labels, which are themselves logically distinct from these criteria; and the view that "real" understanding requires an actual concern for the values in question. The latter ignores the fact that we can understand values we do not share, and which we might be unable to see ourselves ever accepting. The former, wrongly supposing that we could provide purely descriptive criteria for the application of moral terms, and that we can separate attitudes from the features of actions towards which they are directed, ignores the kind of sensitivity and imagination required for even a non-participant understanding of values—of how they can provide reasons for action, of how they are to be applied and developed in particular cases. My own account insists that a concern for *some* values—a participant understanding of them—is essential for understanding *any* values; and that the attitudes, emotions, and values I am to understand must be logically related to those which I share myself: only thus can I come to see, not just *that* others have pro- and anti-attitudes towards actions I can identify by some descriptive criteria, but what these attitudes

amount to, *and* what the actions are towards which they are directed.

A psychopath, who shares in no values himself, is incapable of such understanding, as even detached moral discussion with him will reveal. He does not just fail to appreciate the moral implications of actions which he can independently understand and identify: he cannot even understand what these actions are—what dishonesty, or injury, *are*, as kinds of human action. It is in this sense that he does "not see what they see": he does not see the same actions, any more than an aesthetically insensitive person sees the same pictures as the man who can understand their aesthetic dimensions.

V

However, I must try, albeit briefly, to say more, both about the connection between participation in some values and the capacity for understanding any values, and about the dimension of human life and thought which a psychopath can neither understand nor share.

A psychopath's lack of understanding cannot be limited to some distinct set of concepts, or some distinct aspects of human action, identified as *moral*: in denying the empiricist distinction between fact and value, I am also insisting on the close logical connections between moral and non-moral aspects of action and language; it is not fortuitous that Meursault is as incapable of understanding the emotions and interests of others as he is of understanding the specifically moral features of his actions. These connections can be brought out by showing, firstly, how moral thought and judgment must involve concepts which are not specifically moral, and, secondly, how many of our non-moral concerns involve concepts and attitudes closely related to moral notions.

I argued above that a psychopath cannot understand the nature and quality of his actions, since he cannot identify or understand them in terms of the concepts relevant to moral assessment: but these must include that whole range of concepts, not themselves specifically moral concepts, by which we identify and describe human interests, concerns, and emotions. If we are to understand, for instance, how it is wrong to hurt someone, we must be able to understand what it *is* to hurt someone: and this requires an understanding of the kinds of interest and concern people can have, in the light of which

³¹ C.f. Stephenson, *op. cit.*, ch. II.

actions will be seen as hurtful. We can hurt someone by physical injury; by injuring or insulting someone he loves; by destroying or denigrating his achievements; by ignoring or frustrating his wishes and ambitions; by denying him responses and relationships which matter to him—gratitude, trust, love, friendship. Unless we can understand the significance of such interests, emotions, and relationships in a man's life, we cannot understand what it is to hurt him, or how it can be wrong. An understanding of the moral aspects of my actions, and of the moral values of others, requires an understanding of that dimension of human life which includes both moral values and those interests and emotions which make our actions morally significant: it is this which a psychopath lacks. Meursault was unable to grasp the moral aspects of his conduct *because* he could not understand the meaning of love or friendship, or those concerns of others which reach beyond the immediate moment and its immediate impact on me.

Furthermore, many of the interests, concerns, emotions, and relationships which give moral significance to our actions can themselves be explained only in terms which connect them closely to *moral* concerns. Relationships of love and friendship, for instance, which may play an important part in a man's conception of his own interests, involve a concern for the interests and well-being of the other as having an independent significance, not just as means to some satisfaction of my own. Similarly, many of our ambitions and aims are important to us because we ascribe to the goal or the activity pursued a significance independent of, and indeed basic to, our interest in them. A philosopher may devote time and effort to philosophy not just because he happens to enjoy it, or sees it as an easy way to earn a living, but because he sees the subject and its problems as important—as something he *ought* to be concerned about. Such relationships, to people or activities, are very close to moral relationships: the agent's conception of his own interests as lying in such relationships is based on a *non-self-interested* concern for, and conception of, the significance of someone or something other than himself. And within such relationships a part will be played by more obviously moral notions—of a duty owed, which may conflict with present in-

clination; of the honesty and respect which is required by my concern for the other; of the kind of failure which consists not just in failing to achieve what I want, but in letting the person or the subject down.³²

I am suggesting, not that there is *no* distinction between moral and non-moral values, concerns, and concepts, but that there are close and complex connections between specifically moral concepts and the concepts in terms of which we must explain the interests, concerns, and emotions which matter to people; and thus that a psychopath will be unable to understand not only moral notions, but also these further aspects of the lives of others. He also lacks this dimension to his own life. Meursault not only lacks any understanding of the interests and emotions of others: he exhibits no conception of interests of his own which reach beyond his present moment to other people or other times; no emotional responses or concerns on any level deeper than that of immediate feeling.³³ It is not just that he does not share our particular concerns: he has *none* that we can understand. We cannot say that he is pursuing his own interests, except in the minimal sense that he does what he feels like doing at the time; or that he acts or reacts on the basis of emotions of love, ambition, pride, self-love, or hatred; or that he has any conception of activities, goals, or people as being important to him. He lacks the kind of conception of himself and others which would make this possible.

I have emphasised the absence in a psychopath's own life of that dimension which he cannot understand in the lives of others for two reasons. Firstly, it shows that a psychopath does not have a 'form of life' of his own, radically different from our own, since he has nothing we can identify as a *form* of life. Of course, were we faced by a group of alien beings, in whose lives we could discover as little as we can in the psychopath's, we would not call *them* disordered: we *might* suppose that they have a form of life which we cannot (yet) understand, though we would do better to say nothing at all. But with the psychopath we are not reduced to that despairing silence, which would *not* allow us to call him disordered.³⁴ For he has been brought up within our own culture and language, within forms of life we share: and it makes no sense to suggest that from

³² On these and other features of caring for others, c.f. Milton Mayerhoff, *On Caring* (New York, 1972).

³³ The distinction between emotions and feelings is clearly crucial here: c.f. Sutherland, *op. cit.*, p. 271.

³⁴ This is a further defect in accounts like Fingarette's, and perhaps Haksar's, which seem to allow that the psychopath *has* values, or a form of life, of his own.

such a basis a man could develop a form of life, a set of values and concepts, of his own which have no intelligible connection with those within which he was brought up (just as it would make no sense to suppose that a man could develop his own empirical perspective on, and beliefs about, the world, having no intelligible logical connections with the beliefs and concepts with which, and through which, he first learned to see the world).

The moral rebel develops, extends, and modifies both the values and the concepts of the tradition within which he was brought up: but if his rebellion is to be intelligible as such, there must be a logical continuity between that public, shared, tradition and his present beliefs: we must be able to see (which may require an imaginative effort) how his present values and concepts form an intelligible development of that tradition.³⁵ This is why Camus can make sense of Meursault's life only by showing how it can be understood in terms of concepts and values belonging within an existing, and shared, tradition of thought. But a psychopath does not have his own, revolutionary, values: for he does not share in, and cannot be seen to be modifying, any such tradition, any common form of life. He is *defective*: he has never come to share in a common form of moral or emotional life, and thus can neither understand these dimensions to the lives of others, nor develop them in his own life. His relation to concepts of values and emotion is like that of the intellectually defective to the concepts in terms of which we come to understand and describe the empirical world: having failed to develop a participant understanding of such concepts and language, he cannot be said to have developed his own, different, beliefs and concepts.

Secondly, the absence in a psychopath's own life of any dimension of value and emotion is logically connected to his inability to understand this dimension of the lives of others. We could understand *his* life in terms of values and interests by tracing logical connections between our values and concepts and his: but such connections would also provide the basis for him to understand *us*. If he exhibits some conception of, and concern for, interests of his own, he is in a position to understand

the interests of others, as providing them with reasons for action; if he displays love or friendship for another, he is in a position to understand the significance of such relationships for others, and already has the kind of concern for another which is closely related to a moral concern;³⁶ if he feels resentment, indignation, gratitude, we must ascribe to him a moral understanding of the actions of others, and thus the capacity for a moral understanding of his own actions.

This is not to say that I can understand the values and interests of others only on the basis of a *prior* understanding of values and interests of my own: if that were so, we could imagine someone who acquires a rational conception of his own interests, but never comes to see that there are other people around him with interests and values of their own.³⁷ But one reason for insisting on participation in a *common* form of life, as an essential basis for developing both values and interests of one's own and an understanding of those of others, is to rule out just this kind of possibility. A child does not, and logically could not, come first to feel certain emotions, have certain wants and interests; then learn to describe these in a language; and then discover that there are other people in the world with emotions, wants, and interests like his own. He comes, or learns, to feel emotions, to have wants and interests, in and through coming to share in a form of life with other people—a form of life which includes not just a language, but modes of experience, emotion, and interest. It is by such a process that he comes to be, and to see himself as, a person; to understand and to share in characteristically human activities, experiences, and relationships. And this process is one of coming to see myself as a person among other persons, with whom I share this form of life. An understanding of others, and of the emotions, wants, and interests which inform their lives, is an integral part of coming to understand, and to have, emotions, wants and interests of my own.³⁸

VI

I have been arguing that a psychopath is disordered insofar as he is unable to understand, or to

³⁵ C.f. Beardsmore, *op. cit.*

³⁶ C.f. Bernard Williams, *Morality* (Harmondsworth, 1973), pp. 23–26.

³⁷ This might sound something like the "partial psychopath" defined by Cleckley (*op. cit.*, pp. 195–234), who, unlike the complete psychopath, at least pursues his own interests rationally: I would argue against this that his conception of his own interests will be as limited as his conception of the interests of others; but the question of the nature, possibilities, and limitations of egoism would be a topic for another paper.

³⁸ C.f. Norman, *op. cit.*; P. F. Strawson, *Individuals* (London, 1964), ch. 3.

participate in, a significant dimension of human life and thought, which includes both moral concepts and values and conceptions of self-interest, of emotion, and of concerns other than the strictly moral. Such an understanding, while distinct from a concern for the particular values or interests in question, is not just a matter of an intellectual grasp of the descriptive criteria by which people determine the application of evaluative terms, or identify the objects of their emotions: thus accounts of practical rationality based on the fact-value distinction, in terms of an intellectual understanding to which is added some affective response, are inadequate. A psychopath is seriously defective in practical understanding and rationality: he is cut off by his deficiency from a central dimension of human life, just as the intellectually defective are cut off from the dimension of intellectual understanding and thought. This deficiency is clearly a *disorder*: a psychopath cannot understand the nature and quality of his actions, or the lives and interests of those around him: he cannot control his actions in the light of any rational concerns or values, not because his impulses are strictly irresistible, but because he has no conception of rational values as providing reasons for action.³⁹ We cannot hold him answerable for his actions, any more than we can a young child; we cannot impose on him the duties or expectations we impose on others, allow him the rights we allow others, or hope to share with him the kinds of relationship we share with others: he does not, and cannot, participate in a common life with us.

We can now make more sense of the question of whether a psychopath *could* accept our values, and see the proper basis to the Utilitarian requirement that a responsible agent be open to rational persuasion.⁴⁰ The question, and the requirement, must be taken as logical, not practical. We should ask, not whether the agent will *in fact* change, or be persuaded to change, his values and actions (for that test would be failed by many who we would condemn as wicked, or admire as principled rebels), but rather whether it is *conceivable* that he should come, *rationally*, to accept our values and see his own as wrong; whether we could imagine a rational process of reform, conversion, or moral change,

however unlikely we think it that this will in fact occur. Such a change is conceivable for a person living by intelligible values of his own, who *can* thus understand, whether he bothers to or not, the values and interests in the light of which we assess his actions: for he can understand his actions and our responses to them; he can understand how and why we are blaming and punishing him.⁴¹ But a psychopath is incapable of such understanding: nor does his own life exhibit that basis, of values, concerns, and understanding, from which we could imagine any rational reform developing; and *that* is why he is disordered.

The argument in this paper has been, of necessity, schematic and incomplete: it should provide, at best, the programme for a fuller and more adequate work. Such a work would need to include a more detailed examination of further examples of psychopathy: it would also need to provide a more adequate explication of the account of practical rationality, of moral understanding, and of the connections between moral and non-moral thought and concepts, which is here outlined in the most skeletal of forms. More needs to be said about the role of emotion and emotional capacities in moral development and understanding; about the kinds of logical connections which must be found with, or within, a shared form of life if we are to ascribe intelligible values to a man; about the nature and extent of the connections between moral and non-moral concepts and values. This last point is of particular relevance to the topic of egoism: philosophers have often talked about 'the egoist'—the man for whom only his own interests provide final reasons for action, and whose interests are specified purely in terms of what *he* wants—as if his life provided an intelligible, and even attractive, human possibility; but it would be instructive to compare such a "pure egoist" with a psychopath, as I have described him. We would need to ask what kind of conception of his own interests such a person could have, and what kinds of interest are possible for him; what kind of understanding of other people, their interests, and their values, we can ascribe to him; and how far a life like his is parasitic, conceptually as well as practically, on non-egoistical forms of life.

³⁹ It has sometimes been suggested that the M'Maghten Rules can and should be interpreted in such a way as to cover these kinds of defect of understanding and control: c.f. the *Royal Commission on Capital Punishment* (1949-53), ch. 4; Morris, *op. cit.*

⁴⁰ C.f. above, p. 191.

⁴¹ The moral requirement that a criminal be able to understand that he is being *punished* helps to explain the legal practice of refusing to try, punish, or execute those who have become disordered *since* their offence.

These would be the subjects for a monograph, not an article: I hope at least to have shown that an examination of psychopathy provides a fruitful approach to them: fruitful both for an understanding of mental disorder itself, and for philosophical discussions of practical rationality. For if we can

see that and how a psychopath, as I have described him, is disordered, we will also see the inadequacy of any account of practical reason according to which he is fully competent and rational, and begin to see the outlines of a more adequate account.⁴²

University of Stirling

Received April 13, 1976

⁴² Besides thanking colleagues at Stirling with whom I have discussed the topics of this paper, I must record my special gratitude to Sandra Marshall, whose philosophical co-operation and assistance has been invaluable.

III. ALL ACCEPTABLE GENERALIZATIONS ARE ANALYTIC

HENRY E. KYBURG

1. INTRODUCTION

ACCORDING to my data, the probability is the interval (0.99, 1.0) that between 95% and 98% of introductory philosophy courses mention the statement "All generalizations are false." So far as I know, relatively few people have observed that all general truths are analytic, including that one. (That is, non-observational generalizations: "There are no elephants in this room," need not be taken as analytic.) But it is a thesis that has interesting consequences; among other things it entails that all our fancy theories in physics, biology, chemistry, sociology, psychology, if they are construed as universal generalizations, are completely without content: without observational content, without theoretical content, without any content at all. Furthermore, it enormously simplifies the problem of induction: if generalizations are analytic, their probability is unity, and any evidence whatever will serve to confirm them to the maximum degree. The constraint on explanation, that the theoretical premisses be true, becomes satisfied automatically, since all generalizations and theories will turn out to be true. I am sure one could go on indefinitely citing the interesting and valuable consequences of the thesis that all acceptable generalizations are analytic.

The only cloud that looms on the horizon is the doubt that the thesis is true. (Note that it is not self-certifying in the sense that "All generalizations are false" is self-refuting.) It requires a certain amount of argument to see that it is true; analytic though it be, it is not one of your self-evident analyticities.

Let us make the thesis a bit more precise. Consider any first order language appropriate to a certain body of scientific knowledge. Let a body of knowledge be construed simply as a set of statements in this language. (1) Certain statements come to be accepted into this body of knowledge on the basis of what happens to us. In a certain sense we can *observe* that these statements are true. (2) Certain other statements will become acceptable on the basis of statistical inference. We have

in our body of knowledge a set of data representing observations. These statements support a statistical hypothesis so strongly that it becomes (in a slightly weaker sense) also acceptable. It is statements of these two sorts that I shall take to represent the empirical content of our knowledge. (3) And then there are the scientific statements most people mostly consider: simple universal generalizations. "All samples of table salt are soluble," "Yellow phosphorus melts at 42 degrees centigrade," "All whales are mammals," "All creatures with hearts have kidneys," and of course, "All ravens are black." These statements, I shall maintain, are analytic. (4) Finally, we have the crowning glories of the sciences, theoretical systems: relativity theory, quantum mechanics, genetic theory, and so on. These, especially, I shall maintain are barefacedly analytic.

To defend this thesis, and to show that even if true it does not constitute a massive reduction to triviality of our most cherished beliefs, requires that we attend to each of these kinds of statements in turn, as well as to two other matters. One of these matters is, of course, the notion of analyticity I have in mind. The other is the notion of error. We shall then look briefly at some of the ways in which the thesis might provide a new focus for certain problems in the philosophy of science.

2. OBSERVATION STATEMENTS

It is traditional to divide the vocabulary of the first order language I mentioned into an observational part and a theoretical part, and (usually) to suppose that we can come to know, without inference and with absolute certainty, certain of the statements expressed in the observational language. I shall make an assumption that is somewhat weaker, but roughly within this tradition. I shall assume that within our formal language, we can provide a recursive characterization of a set *O* of statements that are *potentially* capable of being accepted on the basis of observation. I recognize that scientific languages change, and one of the

ways in which they can change is by an alteration in the membership of *O*. Whereas in one language one might simply observe the motion of the sun, in another one might only be able to observe the *apparent* motion of the sun (thereby opening the possibility of the question as to whether or not it *really* moves). It is obvious that we come to accept a lot of the statements of *O*, not on the basis of what happens to us, but on the basis of inference from other statements we accept, on the basis of reliable authority, and so on. Note that *O* itself is the set of all *potentially* acceptable observation statements: thus it contains contraries such as "the next rose I see is red" and "the next rose I see is yellow." The point of singling out this set of statements is to contrast them with statements that we cannot properly be said to know directly on the basis of our experience, such as "Half the draws from this urn yield a black ball," "Three quarters of the peas produced by such and such a cross have red flowers," "All creatures with hearts are creatures with kidneys," "All ravens are black," and " $f = ma$."

In one way I depart from tradition, though perhaps only slightly, and perhaps not from the intent of the founders of the tradition. That is, I shall not assume that when we come to know these statements in *O*, we come to know them with absolute certainty. This relieves us of *some* of the burdens of trying to justify the selection of a particular set of statements to be put in *O*, and also relieves us of some of the traditional problems of epistemology. This is not the place for me to develop or defend a foundationalist epistemology which eschews certainty, but I will mention one typical problem that leads me in this direction. When a person looks at a penny, unless he is a philosopher he will see that it is round. This is certainly true of a naive subject who has no theory of vision, no notion of the rectilinear path of light rays, no projective geometry, to tell him that the pattern on his retina is elliptical. Indeed, were such a subject to claim to see an ellipse, he would be making a mistaken claim; what he *ought* to say he sees, and what (almost) every such person sees, is a round penny. Now of course he may also be mistaken in the traditional way: that is, he may think it is a round penny he is seeing, when in fact it is a trick elliptical penny. In the real world he won't make that mistake very often, but the possibility is there. So philosophers enamoured of certainty, and, more

important, who are in possession of a whole presumably empirical theory of vision, will attempt to save the certainty of observation by saying that he does not *observe* the roundness of the penny, but rather *observes* the elliptical character of its projection, and makes an unconscious (and possibly erroneous) inference to its circularity. We are thus led to replace a simple observation (that the penny is round) with an unconscious observation (that the projection of the penny is elliptical) that is not even accessible to the naive subject until he learns a whole theory of light and vision, and then combine this unconscious observation with an unconscious inference which may not even be truth-preserving! There may be good reasons for doing this (though I don't know them), but it certainly seems worthwhile to start from the outset by accepting the principle that even observation statements may be in error, that we can accept observation statements on the basis of what happens to us, but that we need not suppose that they are incorrigible.

This approach imposes on us an obligation to have a theory of acceptability and a theory of error. We shall consider theories of error in a little while. As for acceptability, we need not go into very great detail, but we should say a few things. I shall take acceptability to be related to *moral certainty* and *practical certainty*, in the following rather technical senses. A statement is morally certain for me, if there is no more than a tiny chance that it will turn out to be in error. What is a tiny chance? That could be taken as a matter of context; I shall suppose that moral certainty can be indexed by a number close to 1, whose exact value can depend on the context. Thus for ordinary purposes I can be morally certain that I am now before my desk; but if the fate of the world were at issue, I might limit my moral certainties to "seeming" statements, and see how strong an argument I can construct on that basis in favor of the proposition that I am before my desk. Practical certainty is a weaker notion: a statement may be practically certain just in case it has a small (as opposed to tiny) chance of being wrong. I suppose that some observation statements are so well supported by what happens to me that they are moral certainties for me. I suppose that other observation statements are well supported, so that I may accept them as practically certain, without their being so well supported as to be morally certain.¹ In addi-

¹ I wonder if the phrase "morally certain" means: it is moral, proper, justifiable, to act as if it were certain?

tion, there are statements so probable, relative to my body of moral certainties, as to be practically certain, even though they are not "observation" statements.

I have already said that the population of O depends on the state of our scientific art, and thus changes through time. But just to give some idea of what, in ordinary contexts at least, we take to be observation statements, they might include such statements as: "There is a table before me," "The penny is round," "The dial reads 10.3," "The current is turned on," "This object is longer than that object," "The scale reads 5.8 grams," "Tom is angry," "There is a cat on the roof," etc. They are to be contrasted with: "The table is a complex structure of cellulose molecules," "The penny is composed mainly of a metal of atomic number 29," "There is a current of 10.3 amperes flowing in the apparatus," "Electrons are flowing through the apparatus," "This object is 3.20 meters long," "This object weighs 5.8 grams," "Tom is suffering from an Oedipus complex," "The cat on the roof is suffering from enteritis."

These statements—the observation statements—all have empirical content in an obvious sense. If I come to accept one of them, I thereby cut down the worlds which are possible relative to what I accept. If I accept it on the basis of what has happened to me, it is the world that has done it: my interaction with the world has caused me to accept it. There are a number of problems here, to be sure, and this whole thesis needs to be provided with some epistemological, and maybe some metaphysical underpinnings. But I assume the general idea is relatively clear for present purposes.

3. STATISTICAL STATEMENTS

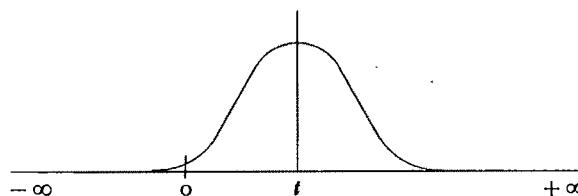
Suppose that I am morally certain that I have examined a sample of 10,000 draws from an urn, and that 5,891 of them have produced black balls. (I could have miscounted, been the victim of hallucination, and so on—but remember we are talking of moral certainty only, not absolute philosophical certainty.) Under certain circumstances the probability that between 0.54 and 0.64 of the balls drawn from the urn will be black is at least 0.99—more properly, the interval (0.99, 1.0). If we take 0.99 as our level of practical certainty, this means that we could simply accept, in our corpus of practical certainties, the statistical statement: the proportion of black draws from this urn lies between 0.54 and 0.64. This is an example

of what I mean by a statistical statement; it is the sort of statement we can accept as a practical certainty; it is a statement having empirical content. Our bodies of knowledge are full of statements like this: "Roughly half the tosses of coins yield heads"; "Close to 51 % of human births are births of males"; and so on.

There are other kinds of statistical statements as well; it is necessary to say something about one other sort, because it will play a large role in our later discussion. Furthermore, although acceptance of statistical statements of this second sort involves no new principles, one principle which we cleverly glossed over in the previous case will come out with clarity and simplicity in this case, and this is a principle which will turn out to have a lot to do with universal generalizations and theories.

In real life statistics the normal distribution is pervasive. This is due to the fact that there are a lot of quantities that are distributed normally; and that in turn is due to a pair of facts: that a lot of things can be construed as additive mixtures of other things; and as a matter of mathematics, almost all additive mixtures of things are roughly normally distributed. In particular, any sum or average of a large number of identically and independently distributed random quantities will be normally distributed, *regardless* of how the random quantities themselves are distributed.

Let us consider the problem of weighing an object on a beam balance of known sensitivity. We assume that we know the sensitivity and that the balance has no systematic error. The readings we obtain will be normally distributed with a mean equal to the true weight of the object, and a standard deviation which is represented by the known sensitivity. As you know, if we looked at the potential infinite set of readings, and plotted them on a graph, knowing the true value of the weight, t , they would look like this:



(This familiar bell-shaped curve is referred to as a curve of error, as well as the normal curve.)

Our problem is that we don't know the true weight of our object, and must obtain that from a

sample of readings. To keep things very simple, we shall suppose our sample consists of a single reading, r . The set of readings, we noted, was distributed with a mean of t , the true value of the weight, and a standard deviation of d , representing the sensitivity of the balance. It follows (purely analytically) that the distribution of the quantity $(t-r)/d$, or error divided by standard deviation, is distributed normally with a mean of 0 and a standard deviation of 1. This is what is tabulated in all those tables at the backs of textbooks in various sciences. The area under the curve is unity (as was the area under the other curve); and the area between two points a and b is the frequency (loosely speaking) with which we will encounter an error less than b and greater than a .

Let us fix on a probability level, say 0.99. There are a number (a very large number: the number of the continuum) of events that have a probability of 0.99 given that appropriate conditions of randomness are met:

The probability is 0.99 that the reading will not be more than 2.5 standard deviations greater than the true value.

The probability is 0.99 that the reading will not be more than 2.5 standard deviations less than the true value.

The probability is 0.99 that the reading will differ by at least 0.01 standard deviations from the true value.

The probability is 0.99 that the reading will differ from the true value by no more than 2.64 standard deviations.

The probability is 0.99 that the reading will either be greater than the true value, or will be at least 2.4 standard deviations less than the true value.

And so on. There are as many such events as there are ways of finding Borel sets on the real line such that the integral of the normal frequency function (corresponding to area under the curve) over the Borel set is 0.99.

Now we weigh our object, and we get a value V . What do we want to say, what *can* we say, about the true value? With probability 0.99 we can say that the true value is less than $V+2.5d$; with the

same probability we can say that the true value falls *outside* the region $V+0.01d$; with the same probability we can say that the true value falls inside the interval $V \pm 2.64d$.

We all know which statement of probability 0.99 we are interested in: the last one. Why? It represents what statisticians call a "shortest" confidence interval. Intuitively, it is because it seems to provide more information than the other statements.² In this case, the most informative statement is the one that assigns the shortest interval to the possible range of weights of the object. In general, we might accept that statistical statement (remember, assigning the weight to an interval is just another way of characterizing a parameter in the distribution of the results of weighings) which has given probability and which assigns the parameter (which may be multidimensional) to the Borel set in parameter space of smallest Lebesgue measure. In any event (jargon aside) we shall suppose we have some way of picking the most informative (complex) statistical hypothesis from among those equally probable, and that this warrants our accepting such statistical hypotheses in our corpus of practical certainties.

All of these things, in my view, are language dependent. Probability is defined relative to a specified formal language; among the specifications of the language is a recursive characterization of certain terms we take to denote potential reference classes, and a recursive characterization of certain functors we take to denote random quantities, i.e., functions from objects to real numbers or vectors. If we are going to have a general theory of scientific inference, we must therefore either have a way of generating an ideal language (not very likely) or a way of choosing between alternative languages in the case in which we have alternatives to choose between (rather more realistic).

4. UNIVERSAL GENERALIZATIONS

So far we have in our rational corpus of practical certainties some statistical hypotheses, e.g., that between 0.49 and 0.51 of coin tosses result in heads; that the distribution of readings of the weight of this object is normal with a mean between $V-2.6d$ and $V+2.6d$ (i.e., the weight of the object

² Of course we cannot *accept* as practically certain all of these statements: in fact they cannot all be true—they are, considered jointly, self contradictory. So if we are going to accept one of the statements (under *ordinary* circumstances: there may be circumstances that call for special treatment, as when we don't care, for some special purpose, whether the object weighs more than so and so much; we just want to be sure that it doesn't weigh less than so and so much), then we should accept the most informative statement.

lies between these limits); that the mature weight of species *X* of frogs is normally distributed with a mean between m_1 and m_2 and a standard deviation between d_1 and d_2 .

Now let us consider how these statements are used. According to my notion of probability, such statements are ingredients in probability assertions. If we know that between 40% and 60% of *A*'s are *B*'s—that is, if the corresponding statement is a member of our set of practical certainties—and if x is a random member of *A* with respect to *B*, relative to what we know, then the probability that x will be a *B* is the interval (0.4, 0.6). Suppose, as is not unreasonable, that what we know is that between 0.99 and 1.0 of the *A*'s are *B*'s, and that we know this at a level of practical certainty of 0.99. Then if x is a random member of *A* with respect to *B*, the probability that it will be a *B* is (0.99, 1.0). This is perfectly adequate grounds simply to *accept* that x is *B* in any lower level rational corpus.

Now suppose that for any term x in the whole language, (1) if " $x \in A$ " is added to our rational corpus, then either: (1) x is a random member of *A* with respect to *B*, relative to that possibly enlarged rational corpus, or (2) it is known already in that rational corpus that x is a member of *B*. If this is so, then for any term x , if " x is a member of *A*" is in our rational corpus, " x is a member of *B*" will be in the rational corpus of any lower level. This is tantamount to either: accepting in the rational corpus of practical certainties, the generalization, "All *A*'s are *B*'s," or (2) adopting as a material rule of inference (a Sellarsian *P*-rule) for getting from statements in the set of practical certainties to statements in any body of beliefs of lower level: From " x is *A*," infer " x is *B*."

It looks as though we are well on our way to a classical reconstruction of scientific inference: we have shown how, in effect at least, a universal generalization can come to be accepted on the basis of enumerative evidence. I used to be more or less persuaded of this, but now I have my doubts. I shall express those doubts in a strong form: the situation described in the preceding paragraphs never happens.

We might say it never happens because all synthetic universal generalizations are false; but this would be a misleading way of putting the matter. A more explicit argument depends on looking more closely at what the observation terms of a language are, and how they are used. Let us suppose first that they are used incorrigibly: then if *S* is an observation statement (belongs to the set

O mentioned earlier), and things happen to our bearer of a rational corpus that lead him to include *S* in his body of moral certainties there's an end of the matter, and *S* is eternally a moral certainty for him. Now if observation statements are used in this way, it is very hard to think of any universal generalizations that don't have exceptions: Bear appearances, for example, are usually, but not always, palpable; they rarely, but sometimes, become person appearances, or disappear altogether. The phenomenal world, if taken seriously, would be a rather spooky place; and the reason it would be spooky is that it is undependable: there are no universal generalizations we can depend on to hold. This is not to say that we could not found our knowledge on statements expressed in such a phenomenological language nor is it to say that we could: that issue may remain open for the purposes of this argument), but simply that all of our knowledge expressed in those terms would be statistical in character, and would not contain universal generalizations even of the sort described above.

Alternatively, suppose that the observation statements are more ordinary ones: "The cat is black," "The scale reads 10.5," "*A* is larger than *B*," and so on. Then it is quite plausible indeed to suppose that we have in our bodies of knowledge universal generalizations relating the terms that appear in these statements: "All samples of table salt are soluble," "All samples of yellow phosphorus under standard conditions melt at 42 degrees C.," "All ravens are black," and so on. But now we can no longer suppose that these terms are used in an error free way. If the right (or wrong) things happen to the bearer of a rational corpus so that he comes to take *S* to be morally certain, it is always possible that other things may happen to him so that he will withdraw *S*, or correct it. Thus if our generalizations concern raven appearances and apparent blackness: we have at best a rough statistical generalization connecting the terms: a few raven appearances have apparent blueness, a few turn out to be colorless, and so on. If our generalization concerns ravens and blackness, then we cannot (to use the immortal words of John Pollock) *know* a raven when we see one. The claim that something is a raven is corrigible. Here is the crux of the element of novelty in my treatment of observation statements: I suppose that they may be corrigible, and at the same time that we do not need *arguments* to support them. They are directly probable.

Now let us suppose that a person has a lot of statements about *A*'s and *B*'s in his stock of moral certainties, and indeed that he has (rationally) accepted in his body of practical certainties the statement that "practically all *A*'s are *B*'s." He cannot accept the universal generalization, since there are also statements in his body of moral certainties to the effect that *this A* is not a *B*, and *that A* is not a *B*, and perhaps a few more. These statements are not incorrigible; we assume that circumstances may arise when they must be rejected. What circumstances? We have a possible candidate right to hand: we are on the verge of a nice universal generalization, which we could accept, were it not for these troublesome counter-examples. To accept the generalization *requires* interpreting the counter-examples as erroneous, and expunging them from our stock of moral certainties. And is this not precisely the hallmark of analyticities?—that in the face of counter-examples we are committed to regarding them as erroneous? So what we can do is make a language shift: regard the terms *A* and *B* to be connected analytically, by virtue of their meanings, come what may (as long as we continue using that language), and to take the judgments we make about *A*'s and *B*'s to be subject to error.

In point of fact, this approach is not so very novel. I have spared you footnotes so far, but it will be worthwhile mentioning here two other writers. Wilfrid Sellars,³ says that one of the things that may happen when we find a lot of Φ that are Ψ , and no exceptions, is to decide to adopt the material move from "*x* is Φ " to "*x* is Ψ " and to reflect this decision by using the modal, " Φ 's are necessarily Ψ 's." "This constitutes, of course, an enrichment of the conceptual meanings of ' Φ ' and ' Ψ '." Robert Binkley,⁴ considers the possible charge of equivocation, if it is argued that when we accept a generalization we change the meanings of the terms involved: the evidence is expressed in old meanings, and the generalization in new meanings. But no equivocation occurs. "When Jones seeks to justify his lawlike statement to Smith . . . he does so by trying to convince Smith that they both have reason to want to change the use of language . . . [to] do as good a job as possible of explaining and predicting what goes on in the world." [p. 73] When one considers that in point of fact when one does accept a generalization, it is often the case

that there *are* (in the old terms) exceptions, the change-of-meaning gambit seems very realistic.

What is lacking in these accounts—and in any accounts that I know of—is a treatment of the arguments that Jones might *rationally* employ to convince Smith that he has a *good reason* to change his language. In the case of theories (as opposed to simple generalizations) Kuhn and Feyerabend seem to deny that there can be rational arguments. We shall return to this issue—and to Kuhn and Feyerabend—in due course.

Meanwhile, when we accept it as analytic that all *A*'s are *B*'s, what has happened to all that evidence we collected? What has happened to the empirical content of our body of knowledge? Should we throw our notebooks away? Surely we cannot regard that evidence as supporting a generalization that is already analytic. No, the evidence is still an important part of our body of knowledge. But rather than supporting the generalization that all *A*'s are *B*'s—which is now without empirical content—it supports a theory of error regarding the terms *A* and *B*. The theory of error tells us that our *A* and *B* judgments are almost always veridical, but sometimes erroneous; and tells us how often. "All *A*'s are *B*'s" has become analytic and incorrigible, but our judgments concerning *A*'s and *B*'s are now corrigible (before they were merely uncertain) and the mass of data we have accumulated tells us how often they are in error. Our body of knowledge as a whole still has the same empirical content; but the empirical content has shifted from the statement "practically all *A*'s are *B*'s" to the statement, "*A* and *B* attributions are practically never in error."

Now this doesn't seem to have gotten us very far: there seems little to choose between a body of knowledge that contains the analytic truth that all ravens are black, and the contingent error statement that practically all things that look like ravens in fact are ravens, and the body of knowledge that contains the statistical hypothesis that practically all ravens are black, and embodies the convention that we can tell a raven when we see one. But if we consider a number of statistical generalizations: almost all ravens have such and such feeding habits; such and such mating habits; such and such nesting habits; and so on, then it may well turn out that the errors associated with these various attributions are not

³ Wilfrid Sellars, *Science Perception and Reality* (London, 1963), p. 357.

⁴ Robert Binkley, "Conceptual Change" in Glenn Pearce and Patrick Maynard (eds.) *Conceptual Change* (Dordrecht 1973).

independent. This allows us to attain a more finely articulated theory of error; and a more informative set of statistical error hypotheses. Thus in principle we can base our choice of language on precisely the same criterion we employed in choosing among alternative statistical hypotheses: choose the one that gives the most information.

A simple numerical example would be very helpful here, but it will have to be extremely artificial. Suppose we have examined a thousand *A*'s, and found that 990 of them are *B*'s, that 992 of them are *C*'s, that 995 of them are *D*'s, that all of them are *E*'s, and so on. By straightforward statistical inference we could infer that almost all *A*'s are *B*'s, that almost all *A*'s are *C*'s, etc. On the other hand, if we took our recognition of *A*'s to be subject to a 1% error (1% of the time when we think we are examining an *A*, in fact it turns out that it is not a "true" *A*), if it happened that in each case it was the same *A*'s that turned out to be recalcitrant, we could accept as analytic that all *A*'s are *B*'s, that all *A*'s are *C*'s, that all *A*'s are *D*'s, that all *A*'s are *E*'s, etc. Of course I am being a little provocative by using the term "analytic," but what I mean is, "necessarily," "irrefutably," . . . you may choose a less provocative word, such as *a priori*. The content that originally lay in the collection of statistical statements, now lies in the statistical statement that we are in error about *A*'s only about 1% of the time. And of course future experience could refute or modify this statement. We could never find an *A* that wasn't a *B*, and a *C*, and a *D*, and an *E*, etc.; but we could easily *suppose* that something was an *A*, and discover, in virtue of the fact that it was not a *B*, that it was not an *A* after all. If this sort of thing happened a lot, we would have to suppose that the frequency of error in making *A* attributions was much higher than the 1% that was originally justified. And if that should happen, we might well be led to suppose that our linguistic "reform" was misguided, and that we would be well advised to return to our former language.

One question remains: the claim that we choose between languages on the same basis as we choose between statistical hypotheses, i.e., on the basis of information content. Roughly speaking we would say something like the following: before the language reform, we had (say) four statistical statements in our body of practical certainties which assigned parameters to the four intervals i_1 , i_2 , i_3 , and i_4 , each including the number 1 as an endpoint. The deficiency of information might be

measured by the sum of the lengths of these intervals. After we accept the universal generalizations, these intervals all become degenerate: the sum of their lengths is 0. But after the language reform we must also take account of the statistical hypothesis regarding errors in making *A*-attributions, and this will give rise to another interval i_5 , in this instance also including 1 as an endpoint. It seems *prima facie* reasonable to compare the length of i_5 to the sum of the lengths of the other intervals. Things are not quite this simple, and this is a complicated matter that would take us too far astray to pursue wholeheartedly; again, you will have to be satisfied with a promissory note. But that is the nature of the subject: epistemology is written on the backs of IOU's.

The obvious objection to all this is that sometimes universal generalizations do get refuted. A classical case in point: the black swans of Australia. Suppose that at a certain point we had accepted it as analytic that all swans are white. How could we then be led to reject that generalization on the basis of the observed black swans of Australia? I ask you to put yourself seriously in the situation of an 18th century Englishman, who *knows* that all swans are white.

A dishevelled sailor, somewhat the worse for drink, whom you meet in a pub, tells you that in Australia he saw some black swans. Do you regard this as refuting the generalization? Hardly. You suppose that he made a mistake, or is lying, or is just telling sea stories, or was tricked by a can of black paint. "All swans are white" remains analytic for you, but your theory of ornithological error has been given a little nudge: it's just possible that the man saw some black birds that he mistook for swans. The theory of error is given a rather larger nudge when the Australian Ambassador tells you, at a formal reception, that there are whole flocks of black swans in Australia. But even this does not (need not, should not) lead you to abandon your universal generalization; you still hold it "come what may," though now your notion of how easy it is to tell a swan when you see one undergoes a radical change; the Ambassador is no ornithologist, but he is probably as good at telling swans as you are. In fact, we may find ourselves in the following situation: we don't know whether to change our language, so that "all swans are white" no longer appears among its analytic truths, or whether to suppose that there is a new species of birds, quite like swans, which makes it difficult to tell when you see a swan. Note

that there is no question of "analyzing" the "concept" of a swan to find an answer: the concept is what we make it, not some entity in a glorious academic mental realm that we can go and look at.

If we are going to maintain that these birds in Australia are swans, then we shall have to change our language so that it is not a mistake to speak of black swans. If we are going to maintain that those birds in Australia are not swans, then we shall have to change our empirical theory of swan-error. The point is that both options are open to us; and another and even more important point is that we are not competent to choose between them: that is a task for an ornithologist. What the ornithologist can do, and we cannot, is to assess the effect on the whole set of ornithological error functions: swan error, gizzard error, feather error, wing-shape error, color error, mating habit error, nesting error, feeding habit error, and so on. Depending on what the observations do to his error functions, he will recommend (and we will adopt) a change of language so that *Swanhoo*d no longer entails *Whiteness*, or he will give the swan-like birds in Australia a name they lacked before.

5. THEORIES

When it comes to theories, the thesis I am proposing is not so novel. Sellars, for example, writes that a number of predicates are implicitly defined if they appear in a set of synthetic sentences "specified as axioms . . . by the *rules of the language* to which they belong." It is true that Sellars calls these sentences "synthetic," but he also says that they are "unconditionally assertable,"⁵ and "assertable by the rules of the language." It is primarily these latter properties that I have in mind when I call generalizations and theories analytic. (There may be a deeper sense of "analytic," but I am unclear as to what it is.) Feyerabend speaks of "different languages and different ideologies."⁶ Kuhn insists that Newtonian "mass" and Einsteinian "mass" "must not be conceived to be the same."⁷ Hockney, whose language is closer to my

own, says that S_1 may be analytic in the language L_1 of theory θ_1 , but false in the language L_2 of theory θ_2 , and "hence, my reasons for rejecting or accepting analytic sentences are just my reasons for accepting or rejecting theories."⁸

Of course this raises a serious problem: what are the reasons, if any, for rejecting or accepting theories? According to Feyerabend "... the freedom of action and thought" of rational people has been considerably restricted by the supposition that there are "certain and infallible rules"⁹ according to which one should decide whether or not to accept a theory. It is a matter, not of argument, but of propaganda, not of logic, but of rhetoric.¹⁰ Kuhn too, seems to regard the fundamental choice between two linguistic frameworks for science as a matter of persuasion. "As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community."¹¹ Sellars, as might be expected, is not so pessimistic: "... the problem concerns the grounds on which a decision to use—that is, to teach ourselves—*this* language rather than *that* can be justified. And to play the language game in which we can be confronted by the need for such a decision, is to know what would constitute a good reason for making it one way rather than another."¹² Thus according to one view, there simply are no "good reasons" for employing one language rather than another, and that is why Science pursues such a jumpy, halting, and irregular course; and according to the other view, all competent scientists (i.e., those who know the rules of the language game) know perfectly well what constitutes such a "good reason," and that is why Science progresses, even if a bit jumpily. The truth, I think, lies somewhere in between. Specifically, I suppose that scientists (and everybody else) has an implicit, inarticulate, and vague intuition of what the "good reasons" in question amount to; but being implicit, inarticulate, and vague, the intuition is not reflected in formal rules of sound argument, but rather becomes a matter of persuasion. (And in persuasion, any-

⁵ Sellars, *op. cit.*, p. 303.

⁶ Paul Feyerabend, "Against Method" in Michael Radner and Stephen Winokur (eds.) *Minnesota Studies in the Philosophy of Science IV* (Minneapolis, 1970), p. 47ff.

⁷ Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1962), pp. 100–101.

⁸ Donald Hockney, "Conceptual Structures" in Pearce and Maynard, *op. cit.*, p. 162.

⁹ Feyerabend, *op. cit.*, p. 101.

¹⁰ *Ibid.*

¹¹ Kuhn, *op. cit.*, p. 93.

¹² Sellars, *op. cit.*, p. 356.

thing goes!) If we can render these reasons articulate, we might be able to perform a useful function.

It is quite true that the classical approaches; verification, confirmation, and falsification, have not turned out well. As Putnam remarks, regarding such statements as " $f = ma$ " and " $e = 1/2mv^2$," "These statements, then, have a kind of preferred status. They can be overthrown, but not by an isolated experiment. They can be overthrown only if someone incorporates principles incompatible with these statements in a successful conceptual system."¹³ As Kuhn puts it, when scientists are confronted by anomalies, they do not "treat [them] as counter instances . . . once it has achieved the status of paradigm, a scientific theory is declared invalid [from the present point of view, an excellent choice of word: not "falsified," not "refuted," but "invalid," a word appropriate to argument forms, to theorems] only if an alternative candidate is available to take its place." And he goes on to remark: "No process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature."¹⁴ It has been pointed out, particularly by Kuhn and Feyerabend, that the data on which two alternative theories with the same subject matter are based is roughly the same: the data that might be said to verify them; the data that might be used to compute degrees of confirmation or corroboration for them; the data which might be alleged to refute one or the other. We are confronted with a problem posed by Quine: ". . . when two theories agree in point of all possible sensory determinants they are in an important sense not two but one. Certainly such theories are, as wholes, empirically equivalent."¹⁵ And indeed this is the typical case for scientific theories: to be sure, not all "observations" are "explained" by each theory; but each theory recognizes that it must come to terms with the observations that exist and no particular observation can be taken as flat out refuting a whole theory.

One response to this situation has been to suppose that scientific terms are simply *vague*. Thus

Scriven, ". . . no exact definition[s] could be adopted which would exhibit significant advantages over the present 'vague' definitions . . . for the crucial terms of theoretical and observational science."¹⁶ Thus Ziff: Our generalizations are hedged about with implicit conditions concerning "ordinary" and "normal" circumstances and objects, whose precise content cannot be unpacked.¹⁷ And Putnam goes so far as to say that even in ichthyology there are no necessary and sufficient conditions for a technical term such as "fish."¹⁸ But this is obviously all wrong; in pursuing science we *do* have necessary and sufficient conditions for the terms we use. (Not, of course, "observationally" necessary and sufficient conditions; that distinction was a clumsy one anyway.) For example, an etymologist on a trip to the Congo can tell when he has discovered a new strain of mosquito (every etymologist must dream of such an occurrence) precisely because a mosquito he finds fails to satisfy the necessary and sufficient conditions for membership in any *known* strain. It is true that *one* specimen won't suffice to convince the scientific community that his is a new strain; one individual may be a sport; from one individual alone it may be hard to decide what the necessary and sufficient conditions for membership in the new strain should be; and thus the successful mosquito catcher must find three or four examples of his new mosquito. But this surely suggests that in the use of scientific terms, at a given time, we *do* employ necessary and sufficient conditions. I say at a given time, for we must recognize that scientific language changes, and I am supposing that it changes almost constantly. Thus it is not surprising that if you look at the behavior of a scientific word over a period of time, you will find that there are no necessary and sufficient conditions for it that have remained constant. But the proper interpretation is not that there are no such conditions, but rather that they have changed with the accumulation of data and the consequent development of theory and acceptance of generalizations.

But if this is so, if scientific language is constantly changing, how can one theory be compared with

¹³ Hilary Putnam, "The Analytic and the Synthetic" Herbert Feigl and Grover Maxwell (eds.) *Minnesota Studies in the Philosophy of Science* III (Minneapolis, 1962) p. 372.

¹⁴ Kuhn, *op. cit.*, p. 77.

¹⁵ W. V. O. Quine, *Word and Object* (Cambridge, Mass., 1960), p. 78.

¹⁶ Michael Scriven, "Definitions, Explanations, and Theories" in Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.) *Minnesota Studies in the Philosophy of Science* II (Minneapolis, 1958), p. 106.

¹⁷ Paul Ziff, "Something about Conceptual Schemes" in Pearce & Maynard, *op. cit.*

¹⁸ Hilary Putnam, "Explanation and Reference" in Pearce and Maynard, *op. cit.*, p. 209.

another? We come up against the problem of sheer incommensurability that has been so lengthily discussed by Feyerabend. If Newtonian mass has one meaning, and Einsteinian mass has another, how can statements about them be said to contradict each other? And if it is to be maintained that this is the case even for terms (such as *raven* and *black*) that are closely tied to observation language, how can we even compare the "all swans are white" theory with the alternative, "not all swans are white?"

Kuhn hints at what I think is the true answer: "... it is precision of articulated paradigm that leads to the recognition of anomaly."¹⁹ This ties in with what I said earlier about error. Every theory—even a common sense theory, implicitly—has associated with it a theory of error, that carries the burden of empirical content associated with the theory. The axioms of the theory are analytic; they "implicitly define" (if we want to use that phrase) the terms that appear in the theory. They thus have no content, so far as the theory is concerned. But there are certain terms that are taken to be directly connected with our experience, not in an incorrigible way, but with a certain implicit chance of error. In advanced sciences, these are measurement terms. We have a theory of "experimental error" which connects the observed magnitudes with true magnitudes. (Note that in exact sciences errors are taken to be distributed normally, which means that *no* observation, however far from the theoretical prediction, could even conceivably *contradict* the theory.) In more primitive and taxonomic sciences, even in common sense knowledge about the world (which I take merely to be a science without a fancy vocabulary), there is a rough and ready theory to the effect that errors don't occur often; but that there are errors is entailed by the existence of generalizations that we take to hold universally. It is to save these generalizations that we introduce misperception, hallucination, inaccurate perceptions, and the like.

All of this may be descriptively accurate, but it doesn't solve the problem we set ourselves: that of articulating the good reasons that lead us to prefer one scientific linguistic framework to another. It does suggest that the place to look is in the theory of error associated with the frameworks. And our earlier suggestion regarding universal generalizations provides a hint as to how to measure the content of the theory of error. Roughly, it is this: that the more closely the theory of error packs the frequency of error to the extremes 0 and 1, the more information carried by our body of knowledge. The language which takes the perception of an elephant to be a necessary and sufficient condition for its existence will have to deal with the fact that there are no useful generalizations about elephants: they come in all colors, they sometimes disappear, they sometimes turn into rocks, and so on. (One might hesitate even to call this a theory—precisely on the ground that it does not seem to have associated with it a theory of error!) A theory which takes elephant perceptions to be sometimes erroneous (and which therefore allows any number of fascinating analytic generalizations about elephants) will have associated with it a theory of error, according to which we are almost never mistaken about elephants under ordinary daylight circumstances, but rather more often mistaken at night, or when we have been drinking heavily. More generally, we might be able to evaluate two theories by looking at the number of statistical generalizations that assign the lower bounds of parameters to a small interval close to one. We might, that is, if we could devise an algorithm for counting numbers of statistical generalizations. I don't know how to do that yet, and if I did this would be neither the time nor the place to do it; but I think there is in this direction some hope of being able to measure the useful empirical content of a theory: i.e., of a set of analyticities plus a body of data, plus statistical generalizations.²⁰

University of Rochester

Received January 15, 1976

¹⁹ Kuhn, *op. cit.*, p. 69.

²⁰ Research leading to this paper has been supported by the National Science Foundation.

IV. THE PROPOSITIONAL STRUCTURE OF PERCEPTION

JOSEPH RUNZO

MY topic is ordinary sense perception. My question is, at least among philosophers, a common one: "What is the structure of the content of perceptual acts?" My aim is not to suggest another complicating factor for this old problem, but rather to offer as spare and simplified an answer as possible.

Analyses of the structure of the content of even our most familiar perceptual acts are often as complex as the acts of perception are themselves commonplace. In particular it is often supposed that there are additional or more directly perceived events or non-physical sensory items which serve as intermediaries between our perceptual acts and the objects or events which we ostensibly perceive. Thus, "sense data," "sense qualities," "appearings," and so on have been postulated. We might refer to such directly perceived intermediary events or non-physical sensory items as "pure presentations." I will argue in this paper that the structure of the content of perceptual acts is simply, and solely, propositional. My argument in defense of this minimalist thesis is twofold. First, this simpler minimalist thesis avoids unwarranted entities which are postulated by pure presentation theories. And second, I will argue that, just where pure presentation theories of perception are seriously inadequate in accounting for certain aspects of perception, this minimalist thesis provides the requisite explanation.

I

It is frequently thought that visual perception involves picture-like mental images and that the counterparts to these mental images are involved in the perceptual acts which correspond to the senses. Thus, on this view, to imagine a scene which one has recently been visually perceiving is to create "in the mind's eye" a picture-like mental image which is similar to but fainter than that in-

volved in the original visual perception. We need to take seriously the resemblance suggested on this "imagistic" theory of perception between the phenomenology of imagining and that of perception. It seems to me that the images involved in imagining *are* very similar to the phenomenological means by which we perceive states of affairs in ordinary perception. But then, considerations applying to the structural form of the one will apply to the same degree to the other. And three objections apply equally against understanding perception or imagining on this "imagistic" model. Moreover, a careful consideration of this natural as well as common imagistic conception of the structure of the content of perception will help expose the fundamental propositional character of perception. For simplicity, let us focus on visual perception and imagining.

(i) When we imagine a face, say, there is no *picture-like* entity which we can *fix* in our attention. Otherwise we could, for an imagined face, inspect myriad and minute details not perceived when we first conjure up the image. But we cannot. The phenomenological content of a mental image during an act of imagining at T_1 has precisely, and only, the features imagined at T_1 .¹

Suppose that at T_1 I imagine Charlie Chaplin's face and describe my mental image to you—straight, dark hair; black, square little moustache. Then you ask, "In your image, was Chaplin's hair parted?" There are two possibilities. Either (a) although I did not describe it, I *was* aware at T_1 either of a part or of the lack of a part in the hair, or (b) it is not the case that there either *was* or *was not* a part in the imagined hair. If I was not aware at T_1 of a part, then subsequently, say at T_2 , to imagine a part in the imagined hair is to imagine a *new* Chaplin-face. Clearly, I think, there would then be two imagined faces; a face imagined at T_1 , which neither has nor lacks a part in the hair, and

¹ Of course, there are facts *about* the mental image which are not part of the phenomenological content of the act of imagining at T_1 . Thus, when I imagine an entity, X , it may be true that I have not imagined Y or Z , and the facts "not εY " and "not εZ " may not be part of the phenomenological content of the act of imagining. But this is a different issue.

a face imagined at T_2 which has a part in the hair. Thus, whatever details of a mental image one is aware of at a particular moment constitute the totality of *that* image imagined at that moment. This is why there are no further details of the "image" of an imagined face which we can peruse, why we cannot scrutinize a mental image in order to discover forgotten facts about an actual face which we are imagining. If mental images cannot be studied in detail, then they are not in any informative sense "like pictures."

Moreover, the notion that mental images are picture-like entities cannot be rescued by claiming that they are pictures which possess only the details which we put in them. For the details of a mental image often have "gaps" and are discontinuous. Thus, I might imagine Chaplin by imagining only his face and lower torso. It may be that I can later "fill in" the missing details of my mental image in order to make the "picture" whole. But *pictures* are not discontinuous in this manner. Even if there is a hole in a painting, photograph, etc., the picture is not simply discontinuous: there is an identifiable *hole* in the picture. So again, since mental images can be, quite simply, discontinuous, they are not in any informative sense "like pictures."

(ii) The fact that images have no details beyond those imagined during a moment of imagining implies that the "imagistic" theory of perception shares an inherent deficiency with the position of the sense-data theorist. A notorious problem for the sense-data theory of perception is the inability to account for the frequent indeterminateness of perception. As a representative of this problem, consider the speckled hen case.² One sees the hen but does not see precisely how many speckles it has. This implies that if perception is to be explicated in terms of sense data, then either those sensory items will have to be indeterminate vis-a-vis speckledness, or if they are determinate in this respect, the percipient must be unaware of that determinate property. These two possibilities present a dilemma because each precludes a salient feature of the hypothesized sense data. Sense data, if they exist, are perceptible entities. Perceptible entities, as opposed to ways of experiencing, etc., are necessarily determinate (though some of the features of percep-

tible entities may not be determinable by certain methods). At the same time, since it is thought that external states of affairs cannot themselves be what the percipient is immediately aware of, sense data, with their determinate features, are introduced into theories of perception to explain *what* percipients are immediately aware of. Thus, since sense data are supposed to be both determinate and perceived in their entirety, the analysis of perception in terms of sense data fails to allow for the fact of indeterminate sense perceptions.

Just as the sense-data theorist cannot account for the indeterminateness of many perceptions, so also if imagining and perceiving are conceived of as a sort of picturing, it is impossible to account for indeterminateness in perceiving or imagining. To take an example, I can imagine the freckled face of Huckleberry Finn without imagining any specific number of freckles. Yet holding *that* "image" of Huckleberry's face firmly in mind, I cannot with careful attention count the number of freckles.³ For if I were to proceed to determine the number of freckles on the imagined face, the image I would *then* be considering would, contrary to the "picture-like" conception of images, not be the same image as the original one. It is not the case that one might be momentarily unable to determine how many freckles are "pictured" by an imagined face. If I imagine Huck's face as vaguely freckled, then it is vaguely freckled. Imagined faces can have determinate or indeterminate numbers of freckles, but then they are different instances of imagining.

Now, if images were like pictures, they would be determinate. For even if the details of a picture *represent* hazy or confused areas, or indeterminate facts, the fine details of the picture itself are never vague, indefinite, indistinct or unspecific. If in a picture, only one side and a portion of another side of a freckled face is pictured, the number of freckles on the half-observed side of the face is indeterminate. But the number of freckle-representing *spots* of paint, charcoal, etc., on the portion of the picture corresponding to the fully-exposed side of the face is determinate. In contrast, the number of spots and the number of freckles on the exposed side of an imagined face can be indeterminate. If images are not determinate, just as if they cannot

² See Roderick Chisholm, "The Problem of the Speckled Hen," *Mind*, vol. 51 (1942), pp. 368-373.

³ Note that an act like imagining a face at T_1 with only three freckles on the nose and three on one cheek—which one could count at T_2 to determine that the face which one imagines has six freckles—is an instance of *initially* imagining a specific number of freckles on the face. Counting those freckles is therefore not a counterinstance here of determining the number of freckles on an imagined face which is imagined at first without a specific number of freckles.

be studies in detail, then with respect to imagining and perceiving, they are not in any informative sense "like pictures."

(iii) Last, there is a vicious infinite regress in the imagistic conception of perception and imagining. For how do we explain, on this account, the means by which the percipient becomes aware of the details of the picture-like image? In terms of another picture-like image? It is readily apparent that the regress that develops here has no virtue, no explanatory value.

And it is not satisfactory to reply that the mind simply "inspects" these hypothesized picture-like images. Why cannot the mind then just attend during perception to the details of the external state of affairs being perceived? Or, for imagining, why cannot the mind then just entertain a set of ideas which expresses the "details" of the image? In both cases the same information will be conveyed as is purportedly conveyed by a picture-like image. In this respect the postulation of a picture-like image does not serve any unfulfilled explanatory function.

Additionally, in the case of perception, scrutinizing the details of picture-like images is *no different in kind* from attending to the details of, for instance, the printed page in front of one: hence the apparent advance in the analysis of perception of having moved from external objects to mental entities is spurious. Picture-like imagistic explanations of perception do not explain the phenomenological means of perception. Such explanations only provide another entity, the "image," whose perception, as much as that of printed pages, etc., needs to be explained.

Drawing all of these considerations about mental images together, it would appear that whatever sort of images might be involved in perception, they are not in any informative sense "like pictures." Yet even this defective imagistic view of perception points to the basic structure of the content of perception. For whether we suppose that perception is imagistic or that there are no intermediary pure presentations and the mind simply attends to the details of the external state of affairs being perceived, awarenesses will be formed of propositions about that external state of affairs. In some sense, then, perception has a propositional character. The question is, in what sense is perception propositional?

II

In order to delineate the propositional character of perception, it will be helpful to distinguish reflective from unreflective perceptual consciousness.⁴ In unreflective perceptual consciousness, the percipient is neither conscious of his awareness of the object of his perception, nor consciously aware of his act of perceiving. Thus for example, when we walk along in a distracted state, a variegated multitude of details—trees, traffic noises, and other immediate objects of perception—are experienced, but unreflectively experienced. In contrast, in a state of reflective consciousness one is *consciously aware of*, reflecting about, one's perceptual experiences. Here, one reflects on an act of unreflective consciousness, for instance becoming aware that *I am perceiving* a tree, etc. We can state the connection between these two types of perceptual consciousness as follows: If a percipient, *P*, perceives an object or state of affairs, *X*, at time *t*, during an act of unreflective consciousness, then in principle *P* could on reflection, during an act of reflective consciousness, become consciously aware that at *t* he is, or was, it would seem to *P*, perceiving *Y*, where *Y* is what *X* looked like to *P* at *t*.

Now, in perceptual acts of unreflective consciousness, it is commonplace *unreflectively* to see a printed page in front of us, to see the black type on the white paper, to hear the long tremolo note of a cello, to smell the freshness of a rose. We can state *what* we perceive when we unreflectively perceive an object or state of affairs by prefacing the description of the content of the perception with "perceive that." We see *that* the printed page is in front of us, *that* its print is black; we hear *that* the note of the cello is long and tremulous; we smell *that* the rose is fresh. In short, in perceptual acts of *unreflective* consciousness, the percipient becomes aware of (what seem to be) facts about states of affairs of his body and environment. What do acts of unreflective perceptual consciousness additionally consist in?

As we have observed, a common answer is that in perception we are additionally immediately aware of pure presentations. Pure presentations are often initially introduced to account for the frequent discrepancies between physical objects as they are and as they appear to us. Sometimes we are fooled by these discrepancies, as when we think that a partially submerged straight stick is bent, or

⁴ These two terms are borrowed from Jean-Paul Sartre's *The Transcendence of the Ego* (New York, 1972).

when, sick and our physiological sense reception processes not working properly, we think that foods are more salty or hot than they actually are. And these discrepancies between the way external states of affairs are and the variant ways they appear, and the consequent perceptual errors which we make, seem to suggest that we are not immediately aware of those external states of affairs. We have already discussed several difficulties with the suggestion that what we are immediately aware of are pure presentations. There is, however, a more telling objection than any yet considered to pure presentation theories of perception: there is a much simpler explanation of the content of our perceptual acts and, consequently, of the variant differences between the way things are and the way they appear.

Consider the fundamental phenomenological facts about perception. First, *phenomenologically* we are unaware of the existence of such pure presentations as sense data—I certainly do not *seem* to perceive, e.g., sense data when I perceive printed pages. Rather, *what* we are, most fundamentally, aware of during perceptual experience are states of affairs of our bodies and environment and (apparent) *facts* about those states of affairs. Given that this *is all* that we are aware of (and would thereby form the basis of the ordinary, unphilosophic notion of perception), considerations of parsimony would argue that this *is all* that constitutes perceptual acts of unreflective consciousness. Second, acts of perception are clearly episodic. One act follows another in quick succession. Third, to perceive is to perceive something *as X*, where *X* is a name or, more commonly, a descriptive locution.

Filling out this third point, let us call a percipient's "sensory field" the "field of view" of some specified set of bodily senses: the physical objects and events which a percipient is at a given moment causally able to detect by means of those senses. Genuine perception minimally involves the differentiation of the perceived entity from its environmental context in the sensory field, even if this act of distinguishing amounts to nothing more than the perception of the entity, whether object or event, as a unified entity (a "something") *in* (i.e. in contrast to) the world. For instance, a percipient can correctly be said to *see* only if what he sees *looks like something*, however amorphous. In short, although to perceive *Y* is not necessarily to perceive *Y as Y*, all perceiving is perceiving *as*.⁵

Since all perceiving is perceiving *as* and conveys (apparent) facts, all perceiving is true or false, or partially true and partially false. If I perceive the paper in front of me as white, then since the paper either is white or is not white, my perception is, correspondingly, either true or false. And since all perceiving is perceiving *as* and always conveys apparent facts which are true or false, the structure of the content of all perception is propositional. This propositional structure points, in turn, to a simple and direct account of the episodic character, the factual content, and the perceiving *as* nature of perception: all perceiving is propositional in structure because to perceive an object or state of affairs, *X*, is, and is no more than, to be episodically aware of a set of *propositions* about *X*.

All that we are aware of in perception is states of affairs of our bodies and environment and (apparent) facts about these states of affairs. But this *prima facie* binary perceptual awareness simply *consists in* the awareness of (apparent) facts about our bodies and environment. To perceive a printed page is to perceive *the page* as well as to perceive that the page is white, in front of us, etc. But it is not as if *the page* were one among several items we become aware of when we perceive a printed page. To "perceive the page" is, and is no more than, a shorthand, if you will, for becoming episodically aware of the total set of (apparent) facts which we do become aware of about the page. This is just what it means to recognize that to perceive *X* is to perceive *X* as something. And thus, a perceptual act of unreflective consciousness is simply an episodically occurrent awareness of propositions about states of affairs of the percipient's body and environment. Or, more concisely, every perceptual act of unreflective consciousness consists entirely in a set of *propositional episodes*. It is simply, and solely, these propositions of which we are episodically aware during perception which convey the (apparent) factual content, and provide the perceiving *as* character, of perception.

Now for example, it is true that the causal history of a given set of propositional episodes will determine, in part, whether or not they constitute a genuine act of perception. If as I sit in Cambridge, Massachusetts, electric probes inserted in my brain cause me to become episodically aware of propositions about the visual appearance of the British countryside, I am not genuinely having visual per-

⁵ G. N. A. Vesey has made the similar, though more restricted, point limited to visual perception that, "All seeing is seeing *as*." (See "Seeing and Seeing *As*" in the *Proceedings of the Aristotelian Society*, vol. 56 [1955-56], pp. 109-124.)

ceptions of Britain. However, the concern here is with what constitutes perceptual acts of unreflective consciousness—viz. sets of propositional episodes. *How* these propositional episodes arise, and what are the exhaustive criteria,⁶ for genuine acts of perception, are separate matters.

At this point one might be puzzled about how, if perceptual acts consist in sets of propositional episodes, one is to distinguish these sets of propositional episodes from each other. For suppose percipient *P* stares at some state of affairs, *S*. It may be very difficult to determine how many discrete sets of propositional episodes occur as *P* stares at *S*. However, this imprecision in the applicability of a propositional episode analysis of perception is no greater than the imprecision in our ordinary notion of what constitutes a *single* act of perception. How many discrete *acts of perception* occur when *P* stares at *S*? There may be clear wrong answers to such a question, but it is not at all clear that there is any uncontroversial right answer. In fact, were the propositional episode analysis any more precise in delineating the number and limits of acts of perception in problematic cases like staring, that would itself suggest that it is not an adequate analysis of our notion of perceptual acts.

This awareness of (sets of) propositions during unreflective perception does not entail thinking about, wondering about, etc., these propositions, though we often do so during the moment of their episodic occurrence. This awareness of (sets of) propositions during perceiving is akin to entertaining propositions, but it neither is necessarily intentional nor necessarily becomes conscious. *Insofar as* it is unintentional and remains unconscious, this awareness is, for example, more like an awareness of propositions during idle daydreaming than like an awareness during imagining or a conscious “entertaining” of propositions. Sometimes intentional and sometimes unintentional, often conscious but frequently unconscious, this awareness is, as far as I can see, a type of awareness of propositions which is unique to perception.

III

The propositional episode analysis of perception has two salient advantages over the pure presenta-

tion analysis. The first advantage is that pure presentations are eliminated and a simpler explanation of the content of perceptual experience is provided. On a pure presentations account, there are two types of entities involved in perception: the pure presentation and the “facts” about or conveyed by that presentation. In eliminating the former type of entity, the propositional episode analysis combines the explanatory function of these two types of entities into the function of the latter entity.

As a set of propositional episodes, each perceptual act of unreflective consciousness concomitantly consists both in the propositional form in which the data of perception is experienced and in the *content*, i.e. the “data,” of that experience. If we think of “presentations” in the broad sense of that “by which we perceive” the objects of perception, then printed pages and roses and musical notes are perceived in acts of unreflective consciousness by means of “presentations.” But there are no perceptions of those presentations: our unreflective perceptual acts *are* those presentations. Perceptual acts of unreflective consciousness have no separable content; they are themselves the contents of ordinary, unreflective perceptual experience. They are both the awareness and the “presentations” of which we are aware in unreflective perception.

Moreover, although this propositional episode analysis is simpler than the pure presentation analysis, it can account both for the differences in the way things are and the variant ways they appear, and for the subsequent phenomena of mistake and illusion in perception, which provided the original reasons for introducing pure perceptions. First, asymmetrical objects appear differently when viewed from several angles. This is simply *because* in each case one becomes occurrently aware of a different set of propositions about that object. There is no need for pure presentations here; there is nothing more to explain about the structure of the content of perception. Second, since sense perceptions, as propositional episodes, are true or false, we can account for perceptual mistake and illusion without postulating a direct awareness of something other than the perceived state of affairs. Thus, suppose I perceive a partially submerged

⁶ A perceptual act consists entirely in a set of propositional episodes, but there are of course additional strictures on what constitutes a genuine act of perception. Suppose someone holds up a card on which are printed sentences about the number 3 and sentences about an apple on the opposite side of the earth. If one reads this card, one will be occurrently aware of a set of propositions about 3 and about the apple. And yet one is not thereby sensibly perceiving either the number 3 or the apple. Thus for instance, as this example shows, one cannot have genuine sense perceptions of *X* unless *X* is sense-perceptible and unless *X* is properly located for one to be sensibly aware of *X*.

straight stick. Quite simply, I might become aware of the proposition that the stick is bent (see it *as* bent), and have a false perception. Or, I might become aware of the proposition that the stick is straight (e.g. see it *as* a straight stick that looks as if it were bent), and have a true perception.

The second advantage of the propositional episode analysis of perception is that it can, as pure presentation theories cannot, account for the common indeterminateness of our perceptions. We have already seen how the sense data theory in particular cannot account for indeterminateness since sense data are themselves determinate. Like sense data, propositional episodes and propositions are themselves determinate. However, propositions can determinately denote vague, indefinite, and unspecific states of affairs; they can ascribe a determinate property to an entity, or they can ascribe an indeterminate property such as "the speckled hen has many speckles." To see the consequences of this for perception, let us go back to the parallel between imagining and perception suggested by the imagistic theory.

On the imagistic theory, when we imagine an object we fix a picture-like image in the "mind's eye." Instead, I suggest that when we imagine an object we become conscious of a set of propositional episodes very like those which, I have argued, occur in ordinary perception. Rather than "reading off" the details of a picture-like image, imagining consists in being episodically aware of just those propositions—excepting those about indeterminate or discontinuous features—which it might be supposed we could "read off" as facts about our mental images, *were* they picture-like entities. Thus, in one instance of imagining Huck Finn's face, there may be a propositional episode the propositional content of which is that Huckleberry's face has many (an indeterminate number of) freckles. And in another instance of imagining Huck Finn's face, there can be a propositional episode the propositional content of which is that Huck's face has, say, six (a specific number of) freckles. In this manner, imaginings *can*, as they in fact *do*, involve vague areas or indeterminate details because they are simply sets of propositional episodes.

Now, like imaginings, ordinary perceptions can

involve perceptions of vague or indeterminate features, or of an absence of certain features of a perceived object. On the propositional episode analysis of perception, which of these occurs will be determined by whether a subset of the causally *possible* set of propositional episodes of a perceptual act is actual and denotes vague states of affairs, or quite simply is absent with respect to certain possible details of the perception. Thus unlike the imagistic theory, the propositional episode analysis of perception can account for indeterminate features and, so to speak, "gaps" in the phenomenological content of ordinary perceptions.⁷

IV

At this point the stalwart proponent of mental images might well try to counter with the charge of methodological equivocation. We certainly *seem* to have picture-like mental images during imagining and visual perception. And after all, a similar phenomenological consideration—namely, that we do not seem to perceive sense data—was counted earlier against the sense data theorist. So, if I am going to give phenomenological considerations some weight, how do I explain the apparent picture-like quality of these apparent mental images? Two considerations can be offered in reply.

First, to answer directly this fair charge, imagining and visual perception *do* involve mental images. All that is denied here is that mental images are *picture-like* entities.⁸ Instead, a mental image is, it seems to me, a logical construct upon the sets of propositional episodes which are acts of imagining or perception. As logical constructs, mental images are only formed during states of reflective consciousness and could not, therefore, constitute the content of perceptual acts of *unreflective* consciousness. As logical constructs, mental images provide unity and continuity to the propositional episodes of imagining and perception. It is this unity and the appearance of continuity provided by the imposition of the mental image logical construct which gives us our false sense that there are further, presently unnoticed, details to our current phenomenological fields which fill out the rest of the "picture."

Second, with scrupulous attention during acts of

⁷ Indeterminateness in sensory perceptions is not a counterexample against my position that all perceiving is perceiving *as*. To perceive vague or indeterminate features of the objects of our perceptions is to perceive them *as*, e.g., colored even though we are uncertain of the exact shade of color, or *as* speckled while the precise number of speckles is not noticed.

⁸ As I read him, Daniel Dennett draws a somewhat similar distinction between images and "pictures" in *Content and Consciousness* (New York, 1969), pp. 135 f., by calling imagining "depictional" or "descriptive" and not "pictorial."

reflective consciousness, we can always become aware of minute details of our *sensory* fields. This misleads us into thinking that all the minute and myriad detail of which we can become aware in our phenomenological field is already there, already "pictured" and awaiting our closer scrutiny. But this is erroneously to assume that a single act of unreflective consciousness could be the object of successive acts of reflective consciousness during which the details of the content of the unreflective act—i.e. the mental image—are progressively more fully noticed.

Consider two acts of reflective consciousness occurring successively at T_1 and T_2 . And consider the unreflective acts, U_1 and U_2 , which are the objects of these two reflective acts and which also occur, respectively, at T_1 and T_2 . Suppose that the details of the unreflective acts, U_1 and U_2 , are similar but also differ. U_2 , the unreflective act which is the object of the second reflective act, occurring at T_2 , cannot be identical to U_1 . All that can distinguish two acts of unreflective consciousness of the same percipient are the content and the temporal span of occurrence of the two acts. To assume that U_1 and U_2 are two moments of the same act because they occur successively is to beg the question; and to assume that U_1 and U_2 are different acts because they occur at two different moments, T_1 and T_2 , is equally to beg the question. On the other hand, *ex hypothesi*, U_1 and U_2 have different contents. Now *some* minimum difference of content must distinguish two successive but similar unreflective perceptual acts, for otherwise no two such acts would ever be distinguishable. And if some difference in content distinguishes two such acts, then, since I do not see how we could meaningfully delineate a requisite minimum difference, *any* difference in content will entail a difference of successive unreflective perceptual acts.⁹ Therefore, new (apparent) details of our sensory fields of which we become successively consciously aware, when we are especially attentive during reflective consciousness, would have to be conveyed in a *new* act of unreflective perceptual consciousness. Hence, during perception we cannot, as we ought to be

able to, successively scrutinize the details of supposedly persisting picture-like images.

In general, any detail of our sensory fields is unperceived if it is not conveyed by one of the actual propositional episodes the complete set of which for any momentary act of perception vis-a-vis a specific sense comprises the totality of the phenomenological field of that sense at that moment. And all causally *possible* details of the phenomenological content of our perceptions are like pains: if we are not (perhaps unconsciously) aware of them, they do not exist. Hence, there is no need to postulate *picture-like* mental images to convey details which we only become aware of later during perception.

V

Two possible objections to this analysis of acts of perception as sets of propositional episodes must now be considered. One objection is that even if this analysis successfully eliminates mental images and other types of pure presentations, another kind of entity, perhaps equally unwarranted, has reared its head. This putatively "simpler" analysis of perception, the objection goes, has eliminated pure presentations at the expense of persisting propositions.

To turn aside one common uneasiness about propositions, we do not, I think, need to regard propositions as Platonic entities which have a reality independent of any saying, thinking, entertaining, etc., of propositions. Perception involves awarenesses of (apparent) facts. These perceptual awarenesses of (apparent) facts about our bodies and environment are analyzable as potentialities for forming dispositions to act. If perception consists entirely in propositional episodes, then our perceptual awarenesses of propositions just are these potentialities for forming act-dispositions.¹⁰ And insofar as our awareness of the propositional content of acts of perception *persists*¹¹—e.g. in the form of acquired beliefs—our perceptual awarenesses of propositions become acquisitions of dispositions to act. Hence, on this account in terms of act-dispositions, the propositions of which we are perceptually

⁹ Thus, for any percipient, P , P 's unreflective perceptual acts, U_1 and U_2 , are different perceptual acts if either (i) there is a temporal moment separating U_1 and U_2 or (ii) U_1 and U_2 are temporally successive but the phenomenological contents of U_1 and U_2 , considered apart from their different moments of occurrence, are not the same.

¹⁰ The fact that many of these potentialities for forming act-dispositions remain unrealized may be one reason that it is often difficult to isolate and identify discrete acts of perception.

¹¹ These awarenesses will both persist and be *conscious* for the percipient when their propositional content becomes part of either the content of a reflective act of consciousness or the content of a later act of memory.

aware need not exist independently of the present or past occurrence of actual perceptions.

However, be this as it may, the more crucial point here is that the content of perceptual acts is delineable solely in terms of its propositional structure, and what view of the nature of propositions is collaterally held in the analysis of perception is ultimately irrelevant to this issue. For the sense data theorist and his cousins will need to hold that on reflection we are aware of facts about the sense data, picture-like mental images, etc., which, on their view, constitute the content of perceptual experience. The form of that reflective awareness is propositional. If the pure presentation theorist conceives of that awareness as involving propositions *qua* Platonic entities, then *he* presupposes *two* types of entities to develop his position: propositions and, e.g., sense data. On the other hand, if he does not conceive of propositions in this way, then the propositional episode analysis of perception can be formulated using whatever other notion of a proposition the pure presentation theorist adopts. Hence, to replace the pure presentation theorist's sense data, mental images, or whatever, the proponent of the propositional episode analysis of perception is not forced to posit a new type of entity—the proposition—with which the pure presentation theorist can dispense. So *whatever* notion of a proposition is employed by the pure presentation perception theorist, the analysis of the content of perceptual acts into propositional episodes provides a simpler explanation of the structure of perception.

A second possible objection against this propositional episode analysis might be that the form of acts of perception cannot be completely propositional. For is there not a perfectly ordinary sense of perception verbs in which they denote purely physical, non-mental, non-conceptual operations? For instance we say, "He must have seen your book. It was right in front of his eyes when he was searching the desk top," meaning nothing more than that, given that his eyes (optic nerves, etc.) were in normal working order and that he was using them, our searcher must have set eyes on the book. *These* acts of perception, it might be objected, are not instances of perceiving *as*; they are not true or false; they do not (necessarily) convey any apparent facts. Hence, they are not propositional in form.

One problem here is that it can be misleading to refer to perception in terms of *acts* of perception. A perceptual act is a peculiar sort of act in that the act is not divisible into something which the agent

does plus the consequential event effected by that action. In this way perceptual acts are unlike such acts as hitting a home run. To hit a home run, at least two events must occur: the player must strike the baseball, and the baseball must, in consequence, be projected over the proper portion of the fence. It may be that, e.g., to see some entity, *X*, I must first direct my eyes in the direction of *X*. But that would at most be a causally necessary condition for the ensuing act of perception and is not part of the *meaning* of *seeing X*. Thus, acts of perception should not be thought of as paralleling duplex acts like hitting home runs. An act of perception consists in only one type of event. It consists simply and solely in the percipient's occurrent awareness of propositions about his body and environment.

Moreover, in the case of visual perception we have another, unambiguous locution to mark usages of "see" to denote a purely physical perception event. We can (really) *see* objects and states of affairs, or only *look at* them. We can look at objects and states of affairs apart from any awareness of propositions about those objects and states of affairs. But merely thus to rest one's eyes on an object is to "see" it only in an extended, parasitic sense. We can meaningfully say, "He's looking at *X*, but he has not seen it yet." What can we possibly mean by such expressions except that to *see X* is to see *X as* something? Thus one sees, for example, Wittgenstein's duck-rabbit line drawing now *as* a set of lines, now *as* a duck, now *as* both a set of lines and a rabbit. Each distinct *genuine* perception is a perception of the drawing *as* something. Or, for instance, consider the parallel for auditory perception exemplified by, "He's listening to the scherzo again, but he has not yet heard the cello passage. (Perhaps he will become aware of it this time.)" Here, although he has listened to the scherzo several times, the listener has not yet discriminated the cello passage *as* something: to perceive *X* is to be aware of a set of propositions about *X*.

To accept the parasitic "looking," "listening," etc., senses of perception as central instances of perception is to think of perception on the two-event model of hitting a home run, and then to regard the first, purely physical, non-mental event as itself a type of perception. Whatever is in our sensory field but is *not* the referent of a set of propositional episodes composing part of the phenomenological content of our perceptual experience, is not actually perceived, even though the percipient may be looking at it, listening for it, etc.

VI

Acts of perception are propositional in form. However, this is not to say that acts of perception are definitely explicable in terms of beliefs. Acts of perception commonly do have a belief aspect, but it is a mistake to identify this, rather than the underlying invariant propositional structure, as definitively constituent of perceptual acts.

Consider the representative analysis of perceptual acts in terms of beliefs of D. M. Armstrong in *A Materialist Theory of the Mind*. Armstrong here defines perception as the acquisition of beliefs, beliefs about the current state of the organism's body and environment.¹² It is part of what we mean by genuine perception that we *tend* (uncritically) to believe what, it seems to us, we perceive. But there are instances of genuine perception apart from belief, and these make it evident that perceptual acts are not definitively occurrent acquisitions of beliefs. Armstrong himself recognizes two such instances.¹³ He treats them as peripheral cases, suggesting that they possibly involve an "inclination to belief" but that they derive their character as genuine perceptual acts from more central cases of perception *with* belief. Armstrong discusses cases in which one's perceptions, such as of mirror images, "do not correspond to physical reality," yet one fails to be deceived by them and hence fails to acquire false beliefs. And Armstrong discusses cases where beliefs are not *acquired* because the percipient already possesses the true (or false) beliefs which would otherwise have been acquired or because the perception is "weak" and ("incredibly," says Armstrong) no belief is acquired.

First, identifying perceptual acts with propositional episodes engenders none of the awkwardness of grouping perceptual acts into central versus peripheral cases which arises from identifying perception with belief acquisition. All acts of genuine perception are, alike, *definitively* occurrent awarenesses of sets of propositions about the percipient's body or environment.

Second, each specific type of perception without belief tells against Armstrong's position. Consider Armstrong's example of the percipient who is not misled by mirror images. For him, the difference between perceiving physical objects by means of

perceiving mirror images and perceiving physical objects under more ordinary circumstances is inconsequential and discountable insofar as both count equally as cases of genuine perception. The additional belief elements in the latter case do not mark a difference between genuine and parasitic cases of perception. In both, though, one definitive aspect making the act a perceptual act is the set of occurrent propositional episodes, whose referent is a state of affairs of the percipient's body or environment. In the mirror image case, some of these propositional episodes simply do not become the subject of belief.

Again, perceptions in which no beliefs are acquired either because the percipient already possesses the true or false beliefs which otherwise would have been acquired, or because the perception was "weak," obviously mitigate against taking the acquisition of beliefs as the definitive characteristic of perception. For example, each successive sense perception of the same entity is a genuine sense perception of that entity even if after the initial perceptual act no further beliefs are acquired, and even if there is no further inclination to acquire beliefs, about the entity in question. If *P* perceives *X* at time *T* during an act of unreflective consciousness, then *P* will have the *potential of forming a disposition* to believe on reflection that he is, or was, perceiving what *X* looked like to him at *T*. But *P* will not necessarily *form* a disposition at *T* to believe on reflection that he is, or was, perceiving what *X* looked like to him at *T*. For if *P* has been continuously perceiving *X* for some length of time previous to *T*, *P* may *already*, before *T*, have formed the disposition to believe that he was perceiving what *X* looked like to him at *T*.

Percipients can have exactly similar perceptions numerous but temporally separated times and they can have exactly similar perceptions in succession. Yet, beliefs are such that a single cognizer cannot acquire a belief which he already possesses. Thus, the acquisition of beliefs cannot be definitive either of similar and separate perceptions, or of successive perceptions, of the same entity. In contrast, percipients can, on numerous occasions, become (perhaps unreflectively) episodically conscious of exactly similar propositions. Therefore, we can

¹² D. M. Armstrong, *A Materialist Theory of the Mind* (London, 1968), p. 209. I do not stipulate that the propositions of which the percipient is episodically aware in perceptual experience are propositions about *current* states of the percipient's body and environment because, for example, a sophisticated astronomer observing the stars might at some moments only become aware of propositions about, and acquire beliefs about, the relative positions of the stars in the distant past.

¹³ *Ibid.*, p. 216 f.

analyze each of a percipient's succeeding similar perceptual acts as a set of propositional episodes which is similar to the set occurring during the first perception of the entity in question.

Moreover, this analysis vitiates the possible suggestion that occurrent beliefs—and not necessarily acquired beliefs—are definitive of perception. If *P* unreflectively stares at *X*, we can explain *P*'s succeeding similar perceptions of *X* in terms of similar sets of propositional episodes. But then, assuming that the notion of occurrent beliefs is intelligible, there is no warrant for explaining *P*'s staring at *X* in terms of succeeding similar occurrent beliefs. For occurrent beliefs would necessarily involve occurrent awarenesses of propositions. Hence, for staring and other perceptions without acquired beliefs, the occurring propositional episodes already explain the structure of the content of perception. The addition of occurrent beliefs has no explanatory value.

It is only because we usually retrospect after forming act-dispositions to believe, that we might think that our unreflective perceptual acts invariably involve beliefs. The propositional episodes which constitute perceptual experience are potentialities to acquire, to the extent to which they persist, dispositions to act in a wide variety of ways. The act-dispositions involved in coming to believe are only one type of these dispositions to act. While acts of perception are, then, definitively neither acquisitions of beliefs nor inclinations to believe, nor even occurrent beliefs, they do have a propositional structure. The data of the phenomenological content of particular perceptual acts are sets of (apparent) facts. These apparent facts can, upon reflection or later recall, become the referents of the

beliefs which are acquired during ordinary reflective perception. Or, they can become the content which is recalled when one remembers "the look," or "the feel," or "the sound," etc., of past perceptual experiences. But the propositional episodes which convey these apparent facts about the percipient's body and environment constitute, in and of themselves, acts of perception.

VII

But isn't this much too arid a notion of perception? For this propositional episode analysis provides no account of the unique "sensory feel" of the contents of perception. Thus, for example, C. I. Lewis even regards the "specific sensuous or feeling character" of the contents of perception as no mere incidental feature, but one of the *criteria* of the "given" in experience.¹⁴ I confess that I am puzzled by this not unlikely line of objection. Like Hume,¹⁵ I can see no difference between our perceptions and our memories or imaginings of our perceptions, beyond the greater "force and vivacity" of the former. If the "sensory feel" of acts of perception is taken as equivalent to the force and vivacity with which they appear in our consciousness, then I do not object to talk about the "sensory feel" of perceptual experience. However, if the "sensory feel" of perceptual experience is supposed to be something else, what can it be? To my mind, the notion of a sensory feel in perception, except insofar as it denotes the force and vivacity of the propositional episodes of perceptual experience, is fictitious. Acts of perception consist solely in episodic awarenesses of sets of propositions about our bodies and environments.¹⁶

The University of Michigan

Received May 6, 1976

¹⁴ C. I. Lewis, *Mind and the World Order* (New York, 1956), p. 66.

¹⁵ David Hume, *A Treatise of Human Nature*, Bk. I, Pt. I, Secs. 1 and 3.

¹⁶ I am indebted to Jack Meiland and Kendall Walton for their helpful comments on an earlier version of this paper.

V. EVENTUAL CHANGE AND ACTION IDENTITY

ROBERT GRIMM

I

FOR a single event to be an instance of more than one kind of action (or a token of more than one action type) it either would have to *be* such at the time of its occurrence, or, being an instance of one kind of action, *come to be* an instance of another (or others). For example, if Daniel Boone shot and killed a bear, his shooting would be his killing only if either it was his killing at the time of its occurrence, or it became his killing. Since his killing awaits the death of the bear, as his shooting does not, his shooting could not simultaneously be his killing. And if events can't change at all, it seems that his shooting can't become also his killing. Other cases than shooting and killing may be less clearly open to the first sort of objection. But this case, at least, suggests that no event can be or become an instance of more than one kind of action, and this has important consequences for the individuation of actions, consequences exploited thoroughly by Alvin Goldman in his work on action.¹

The view that events can't change has had wide and vigorous support. McTaggart apparently thought that the change of events is the only kind of change, and that events change by having been future, becoming present and then past.² But his view provoked thorough criticism and virtually complete rejection. While human beings, animals, trees, and rocks change, events cannot; they just happen or occur. Only that which endures for some stretch of time (continuants) can change.

Events, on the other hand, are only momentary, or perhaps even durationless; they occur, and are gone.³ Further, if events were to change, they would have to have histories longer than their occurrences.⁴ What's more: events themselves are changes, and changes cannot be supposed to change without forcing an infinite hierarchy of time series.⁵ Yet another line of attack appeals to a distinction of the spurious from the genuine. Futurity, presentness, and pastness are not genuine characteristics, hence events cannot lose or acquire them.⁶ Or more generally, some changes are genuine and some are not, and reputed changes in events fall into the latter category.

I shall argue that an event such as Boone's shooting does become his killing, that, in general, events can *come to be* instances of more than one action kind.⁷ I shall comment on the various objections to eventual change mentioned above and examine the arguments of Goldman and others for the distinctness of actions such as Boone's shooting and his killing. My claim is that eventual changes allow for the identity of actions such as these. Whether these eventual changes are *genuine* or not, as it turns out, is beside the point.

II

Let us begin with an argument to show that events can and do change, and in particular that they change by becoming causes after they've already occurred. It seems to me that any relational characteristic of non-simultaneous events

¹ Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, 1970).

² J. M. E. McTaggart, *The Nature of Existence* (Cambridge, 1927), Volume II, Book V, Chapter 33.

³ G. E. Moore, *The Commonplace Book of G. E. Moore*, ed. by C. Casimir Lewy (London, 1962), Notebook II (c. 1926), entry 8, p. 97.

⁴ Moore again, *ibid.* C. D. Broad, *Examination of McTaggart's Philosophy* (Cambridge, 1938), Volume II, Chapter XXXV, pp. 277-281.

⁵ Broad again, *ibid.*

⁶ J. J. C. Smart, "The River of Time" in *Essays in Conceptual Analysis*, ed. by A. G. N. Flew (London, 1956), pp. 213-227, see p. 224.

⁷ J. J. Thomson discounts this thesis in "The Time of a Killing," *The Journal of Philosophy*, vol. 68 (1971), pp. 115-132, see p. 132. Jonathan Bennett has argued for it in "Shooting, Killing, and Dying," *Canadian Journal of Philosophy*, vol. 2 (1973), pp. 315-325. Since the completion of this paper, it has been brought to my attention that John F. Vollrath also has argued for it in "When Actions are Causes," *Philosophical Studies*, vol. 27 (1975), pp. 329-339 and Irving Thalberg against it in "When do Causes Take Effect?," *Mind*, vol. 84 (1975), pp. 583-589.

will do the trick, but causation is especially interesting and relevant in connection with my thesis about actions. The fact that I'm appealing to relational characteristics particularly lays my conclusion open to the charge of spurious change, or "Cambridge change," as Geach has come to call it,⁸ but I shall remark later on how and why this charge misses the mark in connection with the problem of action identity.

The argument to show that events change by becoming causes follows:

- (1) Causes and effects are events.
- (2) Events occur in time (i.e., at moments, t_i 's, of time).
- (3) Causes temporarily precede their effects.
- (4) *Prima facie*, every singular statement attributes a characteristic to its logical subject. If the statement is true, the logical subject has the characteristic. If not true, it doesn't have it.
- (5) If something does not have a certain characteristic at some moment of time t_i , and does have that characteristic at some later moment of time t_j , then that something has *changed*, in that it has acquired that characteristic.
- (6) The truth of a singular causal statement, such as " A_1 caused B_1 " (where A_1 and B_1 are particular events of kind A and kind B respectively), is contingent just upon:
 - (i) there being a causal law which "covers" what is stated, and
 - (ii) the occurrence of *both* A_1 and B_1 .
- (7) If A_1 occurs at t_i , and B_1 occurs at t_j ($i \neq j$, and j later than i), then at t_i it is not true that A_1 caused B_1 (even if " A 's cause B 's" is a causal law) (by (6(ii))).
- (8) If A_1 occurs at t_i , and B_1 occurs at t_j , and there is a causal law which "covers" their occurrences (such as that A 's cause B 's), then at t_i it is true that A_1 caused B_1 (by (6(i)(ii))).
- (9) If A_1 occurs at t_i , and at t_i it is not true that A_1 caused B_1 , then at t_i A_1 does not have the characteristic of having caused B_1 (by (4)).

(10) If A_1 occurs at t_i , and B_1 occurs at t_j , and it is true at t_i that A_1 caused B_1 , then at t_j A_1 does have the characteristic of having caused B_1 (by (4)).

(11) If at t_i A_1 does not have the characteristic of having caused B_1 , and at t_j A_1 does have this characteristic, then A_1 has *changed*, in that it has acquired that characteristic (by (5)).

(1) has been disputed,⁹ but it's plausible and I shall assume that it's true. (2) needs no argument. (3) may not be universally true—some causes and their effects may be simultaneous, and it may even be possible for effects to precede their causes—but all that is needed here is that at least some causes precede their effects. Against (4) it can be claimed that some singular statements do not attribute characteristics to their logical subjects, or that they don't attribute genuine characteristics. I am supposing that any open sentence is associated with a genuine characteristic unless there are good reasons for disallowing it. Relational statements especially have been suspect on this score. I shall say more about this later. (5) doesn't give an exhaustive characterization of change, it only specifies one broad kind of change. (5), in fact, is a limited version of what Geach has dubbed the "Cambridge criterion" of change.¹⁰ He rejects it on the grounds that it countenances changes which are not "real" changes, changes which he calls "Cambridge changes." He offers intuitive counterexamples to the criterion, but admits that he has no more refined criterion which would capture the real/Cambridge change distinction. There is a problem here, but it doesn't touch the issue of action identity.

(6)(i) follows the Hume-Mill tradition in holding that a singular causal statement entails that there is a covering law. But it also accords with Ducasse's criticism of that tradition to the effect that a singular causal statement does not entail some specific causal law.¹¹ Here I am persuaded by Davidson's reasoning concerning the consistency of the two accounts, due to the distinction between knowing there is a law covering two events and knowing what the law is.¹²

⁸ P. T. Geach, *God and the Soul* (New York, 1969), pp. 71–72.

⁹ For instance, see Zeno Vendler, "Effects, Results and Consequences" in *Analytic Philosophy*, ed. by R. J. Butler (New York, 1962), pp. 1–15, and "Causal Relations," *The Journal of Philosophy*, vol. 64 (1967), pp. 704–713.

¹⁰ Geach, *op. cit.*

¹¹ C. J. Ducasse, "Critique of Hume's Conception of Causality," *The Journal of Philosophy*, vol. 63 (1966), pp. 141–148.

¹² Donald Davidson, "Causal Relations," *The Journal of Philosophy*, vol. 64 (1967), pp. 691–703, see pp. 701–702.

(6)(ii) is essential to the argument. Its not the case that A_1 caused B_1 before B_1 occurred. On the basis of a law to the effect that A 's cause B 's, one could infer that, given the occurrence of a particular A -event (A_1), a particular B -event *will* occur (or probably will occur, depending on the nature and status of the law). But this is just to say that the particular A -event *will* cause, or will be the cause of a particular B -event. So it's not just that A_1 's being the cause of B_1 is *discovered* when B_1 occurs, for " A_1 caused B_1 " *entails* that B_1 occurs or has occurred. The *truth* of " A_1 caused B_1 " is at stake on the occurrence of B_1 , not just the discovery of that truth.

The question can be raised as to whether A_1 's *causing* B_1 is itself an event, over and above the events A_1 and B_1 . If it is, one may ask when *that* event occurred. On the other hand, if it's not, then although there is a time at which " A_1 caused B_1 " is true (i.e., " A_1 caused B_1 at t_1 " is true), there is no event of A_1 's causing B_1 which occurs at that time. If it's supposed that A_1 's causing B_1 is itself an event, then how is it related to A_1 and B_1 ? Does it itself have causes and effects? If so, what are they? For example, suppose an explosion causes a window to break. Does the explosion cause the explosion's causing the window's breaking? And does the explosion's causing the window's breaking cause the window's breaking? Affirmative answers lead to regress. But negative answers leave the question of what *are* the causes and effects of causings? Whichever way we turn, it looks like either the causing is just the explosion or the window's breaking or we're off and away on another regress. I conclude that A_1 's causing B_1 is *not* itself an event over and above A_1 and B_1 . Still, there is a time at which it's true that A_1 causes B_1 , and that's all that's needed for the argument. With these remarks to bolster (1) through (6)(i)(ii) in the argument, I claim that the rest follows—that events change by becoming causes.

III

Now for some remarks on the objections to eventual change. The contrast of continuants with events does little but mark a distinction. The question is, *why* should the fact that human beings, animals, and things endure for some stretch of time while events do not, allow changes of the former, but not of the latter? The basis for this

thought appears to be the notion that events are momentary, so that the minimal time required for *any* change to take place is longer than the maximal time involved in the occurrence of any event. Ordinary events, such as explosions, battles, shootings, and so on, do not appear to be momentary but to take time in their occurrence. Consequently, it's easier to see how events which *take* time can change in the course of their occurrences. For instance, the melting of a piece of butter can change in that the rate of melting will increase or decrease proportionally to the amount of heat being applied. But perhaps a distinction is needed between processes and events.¹³ In any case, momentary events constitute the hard case for eventual change, so I'll adhere to the notion that events are momentary. The supposition is, that the duration of the occurrences of momentary events is too short for them to undergo any changes. An argument for the point might go like this: If events do not occur through time, they at least occur at a time. Suppose then that an event e occurs at t_1 . Suppose further that consideration is given only to changes in which either something first has a certain characteristic and then doesn't have it, or first doesn't have it and then has it. For something to change, then, there must be a time at which it has (or doesn't have) the characteristic, and a time at which it doesn't have (or has) the characteristic. So, for something to be able to change, it must endure through at least *two* units of time. An event such as e occurs at a *single* unit of time. Therefore, it doesn't last long enough to undergo a change during its occurrence. So if anything which changes must change during its existence or its occurrence, then events can't change. This also explains why, if events were to change, they would have to have histories longer than their occurrences. And alternatively, it shows why changes in events would force an infinite hierarchy of time series. For if it is insisted that any changes in events *must* take place during their occurrences, then there would have to be another time series within which events have durations of occurrence of at least two units of time. An event e' , which was a change in the event e , if it were supposed to change during its occurrence, would require yet another time series, and so on *ad infinitum*. The upshot of these objections, it seems to me, is that events cannot change during their occurrences, so that if they do change, they must change either before or after their occur-

¹³ A. N. Prior, *Time and Tense*, I. "Changes in Events and Change in Things" (Oxford, 1968), pp. 1-14, see pp. 2-3.

rences, and thereby must somehow have histories longer than their occurrences. The argument (1)–(11) shows that events change by becoming causes. So it appears that we must accept the notion that events do somehow have histories longer than their occurrences. Despite its apparent strangeness, this result is not limited to events. A man may become a grandfather (or even a father) after he already has died. Once again, of course, the case turns on relational characteristics. But this shows that such changes are not limited to events. Continuants also change in this way.

And so I turn to the notion that either being the cause of x or being the grandfather of x are not genuine characteristics, or that such changes are not genuine changes but perhaps only "Cambridge changes." The basic issue here is how such characteristics or such changes would affect the identity of those items which supposedly have them or undergo them. The answer is that if the characteristics or the changes are genuine, then they don't affect the identities of the items which have them or undergo them. For genuine changes require that that which changes be the same before and after the change. And the same point holds even more clearly if the changes are not genuine. Thus, the question of whether being the cause of x is a genuine characteristic, and the question of whether an event's changing by becoming a cause is a "real" change, are irrelevant to the issue of whether or not single events can be or become instances of multiple action kinds.

Suppose Boone shot the bear at t_1 and Boone's shooting the bear caused the bear's death at t_1 , so that Boone's shooting the bear, which occurred at t_1 , became the cause of the bear's death at t_1 . Boone's shooting the bear caused the bear's death *if and only if* Boone caused the bear's death by shooting him. And Boone caused the bear's death by shooting him *if and only if* Boone killed the bear by shooting him. So Boone's shooting the bear caused the bear's death *if and only if* Boone killed the bear by shooting him. And insofar as Boone's shooting the bear became the cause of the bear's death at t_1 it became Boone's killing the bear by shooting him at t_1 .

But, it is *neither* the case that Boone killed the bear by shooting him *if and only if* Boone killed the bear, *nor* that Boone caused the bear's death *if and only if* Boone killed the bear. Concerning the first:

Boone could kill the bear in some way other than by shooting him. Boone could stab the bear. As for the second: Boone could cause the bear's death in some way other than one of the ways in which he would have killed him. Boone could inadvertently frighten someone else who has rifle in hand, thereby causing him to fire his rifle, with the rifle ball striking the bear, who thereby dies. Boone would have caused the bear's death without having killed him. But in the case under consideration, Boone caused the bear's death by shooting him at t_1 , so Boone killed him at t_1 . And since Boone caused the bear's death by shooting him at t_1 , Boone killed the bear by shooting him at t_1 . So, in this case, that Boone killed the bear at t_1 comes to the same thing as that Boone caused the bear's death by shooting him at t_1 . Boone killed the bear *if and only if* Boone caused the bear's death by doing something or other of a kind appropriate to killing. In this case he shot him. Thus, since Boone's shooting the bear became the cause of the bear's death at t_1 , it became Boone's killing the bear at t_1 .

If one claims that causing the death of the bear is not a genuine characteristic, or that Boone's shooting the bear's becoming the cause of the death of the bear is not a "real" change, then either one must also claim that Boone's killing the bear is not a genuine or "real" action, or else admit that it is, despite the spurious nature of the characteristic or the change. By anyone's intuitions I expect that Boone's killing the bear counts as a real action. So I set aside the objections from the distinctions between genuine and spurious characteristics and real and Cambridge changes.

IV

Against the background of this account of events becoming causes, and of shootings becoming killings, let us look at the arguments which have been used by Goldman, Thomson, and Davis against the identity theory for actions. My concern is not to show that actions in the means relation always are identical, but rather that they can be. This is enough to show that Goldman's identity conditions for actions are much too stringent.

A survey of the arguments against the identity theories of Anscombe and of Davidson¹⁴ in Goldman's book and in Thomson's and Davis's articles shows that they are mostly of the following form:

¹⁴ G. E. M. Anscombe, *Intention* (Oxford, 1958). Donald Davidson, "Actions, Reasons, and Causes," *The Journal of Philosophy*, vol. 60 (1963), pp. 685–700, and elsewhere.

For any items a and b , if a is identical with b , then whatever is true of a is true of b and conversely, so, if something is true of a and not of b , or conversely, then a is not identical with b . Goldman invokes differences in the effects or consequences of actions, differences in the causes or enablers of actions, and differences in the supererogatoriness of actions. He also points to the asymmetry and irreflexivity of what he calls "the by-relation," indicating that items for which any asymmetrical and irreflexive relations hold cannot be identical. And he suggests that the identity thesis is objectionable because it requires rejection of the basic-nonbasic action distinction.¹⁵ This last is really subsidiary to the argument from "the by-relation." For Goldman bases the distinction on "the by-relation."

The line of argument of Thomson and of Davis is that since actions in the means relation occur at different times, they cannot be identical.¹⁶ Thomson presents three special problems which arise on this basis for the identity thesis. She calls them *the tense problem*, *the date problem*, and *the temporal order problem*. The tense problem has actions such as shooting and killing respectively having and lacking the property of having occurred (e.g., before the person shot dies). The date problem has them respectively having and lacking the property of having occurred at a certain time t_0 . The temporal order problem has such actions respectively having and lacking the property of being followed n hours later by someone's death.

One way out of these difficulties would involve holding that the troublesome contexts, e.g., temporal contexts, are all non-extensional. Goldman mentions this, but suggests that the requisite proof of opacity would involve begging the question. Another way out, also mentioned by Goldman, is to claim that properties hold of actions only under specified descriptions. His response is that this requires unnatural ways of speaking about actions, and would commit us to the objectionable view that causation and the having or lacking of other properties by actions are somehow language-dependent.

I suggest, however, that all of these problems are avoided if events change by acquiring action-kind characteristics, as Boone's shooting the bear becomes his killing the bear with the death of the

bear. Boone's shooting the bear may have different effects or consequences from Boone's killing the bear. They may also have different causes or enablers. And the event which is Boone's shooting the bear acquires the characteristic of being Boone's killing the bear at a different time from when it is Boone's shooting the bear. But none of this requires that it is not the same event which has both characteristics.

V

The supposition that it does rests on the presumption that the identity of actions must satisfy Leibniz's Principle of Identity of Indiscernibles. But even the identity of continuants does not satisfy this principle. A leaf before and after its color and shape have changed is yet the same leaf, though the leaf before and the leaf after the change do not have all properties in common. The general point is that the *P.I.I.* is inconsistent with change, and yet change presupposes identity through change.

Once again, the general form of the arguments against the identity theory of actions is that if $F(a)$ and $\sim F(b)$, then $a \neq b$. This sort of argument is valid just in case " $F(a)$," " $\sim F(b)$," and " $a=b$ " are mutually inconsistent. There is a number of possible responses to this alleged inconsistency. I count five. I'll list them and comment on each.

(1) " $F(a)$," " $\sim F(b)$," and " $a=b$ " are inconsistent, so, given $F(a)$ and $\sim F(b)$, one validly concludes $a \neq b$. This is a valid form of argument, I suggest, just insofar as a and b are space-time (or atemporal) individuals. For the *P.I.I.* can only hold for changeless individuals, not for individuals which change. This is shown by the fact that given the *P.I.I.*, $F(a)$ and $\sim F(a)$ (where the individual a changes by losing the property F), yield $a \neq a$.

(2) " F _" is a non-extensional context, such as "Jones believes that $G(a)$ " (where " F _" is "Jones believes that G "). Then " $F(a)$," " $\sim F(b)$," and " $a=b$ " are not inconsistent, and given $F(a)$ and $\sim F(b)$, one cannot conclude that $a \neq b$. Indeed, a standard test for the non-extensionality of a context such as " F _" is whether from " $F(a)$ " and " $a=b$," one validly can conclude " $F(b)$." If one's ground for holding that " F _" is non-extensional is the failure of substitutivity for " F _" then the individuals a and b again must be spacetime (or atemporal) individuals. For one could only draw this conclusion if one had good

¹⁵ Goldman, *op. cit.*, pp. 2-6.

¹⁶ L. H. Davis, "Individuation of Actions," *The Journal of Philosophy*, vol. 67 (1970), pp. 520-530. J. J. Thomson "The Time of a Killing," *The Journal of Philosophy*, vol. 68 (1971), pp. 115-132.

reasons for thinking that a and b satisfy the *P.I.I.*, and had some sort of independent grounds for accepting the truth of $a=b$. (Otherwise, for example, one might conclude that the substitution doesn't go through because of change.)

(3) The F which a has and the F which b lacks are not the same. Suppose " $F(a)$ " and " $\sim F(b)$ " filled out with temporal indices, with those indices construed as attaching to the predicates, so that the predicates in question are " $F-at-t_1$ " and " $F-at-t_2$." " $F-at-t_1(a)$," " $\sim F-at-t_1(b)$," and " $a=b$ " are not inconsistent, so the first two don't yield $a \neq b$. Since this preserves the notion that a and b satisfy the *P.I.I.*, it once again presumes that a and b are space-time (or atemporal) individuals.

(4) a and b are distinct, but distinct temporal slices, stages, or phases, of the same space-time individual. Once again, suppose " $F(a)$ " and " $\sim F(b)$ " filled out with temporal indices, but now construe the indices as attaching to, or being built into, the singular terms. Let A be a space-time individual, such that $a=A-at-t_1$ and $b=A-at-t_2$. Then " $F(a)$ " is " $F(A-at-t_1)$ " and " $\sim F(b)$ " is " $\sim F(A-at-t_2)$." One may still hold that there is a relation of kinship between a and b which reflects the fact that, though distinct, they are temporal slices of the same space-time individual.¹⁷ Neither space-time individuals nor temporal slices of space-time individuals undergo change. So this move once again preserves the *P.I.I.* by restricting the domain of individuals for its interpretation to changeless individuals.

For individuals a and b which can change, I conclude that all four of these responses to the alleged inconsistency of " $F(a)$," " $\sim F(b)$ " and " $a=b$ " are inappropriate. For they all hew to the *P.I.I.*, which is inconsistent with change. Continuants, such as Boone and the bear, certainly can change. And I've argued that events, such as an explosion or Boone's shooting a bear, can change. So if a and b are individual continuants or individual events, none of the above four responses is appropriate. It occurs to me that perhaps the notion of a temporal slice individual is the model for the notion of unchangeable events. Its easy to see how, if this is how events are conceived, they should be thought to be unchangeable. The trouble is that ordinary events, such as the explosion of a bomb or Boone's shooting of a bear, "involve" continuants, such as the bomb, Boone, and the

bear. But temporal slice individuals cannot "involve" continuants.

If these four responses must be abjured for changeable individuals, then the only option left is to give up or appropriately modify the *P.I.I.* in a way which renders it consistent with change. The kind of modification which seems appropriate involves temporalizing the *P.I.I.*, i.e., introducing a temporal variable. This can be done in a couple of different ways, the difference turning on the scope of the temporal quantifier. I shall list them under option (5), and comment on each.

(5)(i) $(\forall x)(\forall y)(\forall t)(I(x, y, t) \supset (\forall F)(F(x, t) \equiv F(y, t)))$ ¹⁸

This temporally modified version of the *P.I.I.* is unsatisfactory for the following reason: Given a t_1 and a t_2 ($i \neq j$), " $I(a, b, t_1)$ " is consistent with " $\sim I(a, b, t_2)$." For " $(\forall F)(F(a, t_1) \equiv F(b, t_1))$ " is consistent with " $(\exists F)(F(a, t_1) \& \sim F(b, t_2))$ ". In other words, a could be identical with b at one time, and not at another. This might have some sense if interpreted in terms of alternative possible worlds, but even if it does, its much more than we need or want to claim for changeable individuals. This suggests that identity itself should not be relativized to time, but that temporalization should be restricted to the right-hand side.

(5)(ii) $(\forall x)(\forall y)(I(x, y) \supset (\forall F)(\forall t)(F(x, t) \equiv F(y, t)))$ ¹⁹
T.P.I.I.

Since identity is not relativized to times on this version, the previous result is ruled out. Yet the *T.P.I.I.* does allow that there should be a t_1 , a t_2 and an F , such that $F(a, t_1)$, $\sim F(b, t_2)$, and yet $a=b$. In other words, the *T.P.I.I.* is consistent with change, as the *P.I.I.* is not.

But for the *T.P.I.I.* not to reduce to the *P.I.I.*, " $F(a, t_1)$," for instance, must *not* be construed either as " $F-at-t_1(a)$ " or as " $F(a-at-t_1)$." For, so construing it, we would have returned either to option (3) or to option (4), and hence, the *T.P.I.I.* would still be inconsistent with change. Consequently, " $F(a, t_1)$ " must be construed in some other way. I propose to take the temporal indicator as attaching to the statement as a whole. The three cases can be distinguished in the following way: For option (3), " at^1-t " is an operator on predicates which generates predicates. Thus, " at^1-t " attached to the predicate " F " yields the predicate " $at^1-t(F)$." Similarly, for option (4), " at^2-t " is an operator on singular terms which generates

¹⁷ E.g., cf. W. V. O. Quine, "Identity, Ostension, and Hypothesis" in *From a Logical Point of View* (Cambridge, 1953), pp. 65-79.

¹⁸ Compare E. J. Lemmon, "Comments on D. Davidson's 'The Logical Form of Action Sentences'" in *The Logic of Decision and Action*, ed. by N. Rescher (Pittsburgh, 1966), pp. 96-103, see p. 100.

¹⁹ See Nicholas Rescher and Alasdair Urquhart, *Temporal Logic* (New York, 1971), p. 242.

singular terms. Finally, " $at^3 - t$ " is an operator on statements which generates statements. The choice among these operators, then, determines whether or not the *T.P.I.I.* is distinct from the *P.I.I.* and is consistent with change. Only the selection of " $at^3 - t$," the statement operator, will do the job. So, in the *T.P.I.I.* I construe the temporal indicators in this way. An operator such as " $at^3 - t$ " may be combined either with tensed or untensed statements.²⁰ But I shall restrict myself to untensed statements. In addition to the adoption of the *T.P.I.I.* so construed, some comments are required on the notion of the happening or occurring of events, before I can return to the arguments against the identity theory of actions.

Events occur or happen; things do not. Things change, events happen. This was Moore's view of events. "It certainly can't be, as language suggests, that the same event *is* at all times, & possesses at one a characteristic wh. it doesn't possess at others. This would assimilate an event to a thing wh. persists & has at one time a quality wh. it hasn't got at others. You get a definite contradiction: If you suppose an event *is* at any other time except that at which it is present: that at wh. it is present *means* the time at wh. it *is*. How can an event have a characteristic at a time at wh. it isn't?"²¹

Rather than the occurrence of events being the counterpart of the change of things. I suggest that it is the counterpart of the existence of things. A bomb exists. An explosion occurs. And just as existence is not a property of the bomb, so occurrence is not a property of the explosion. As "a bomb exists" is taken to mean "there is a bomb," "something is a bomb," (or " $(\exists x)(\text{Bomb}(x))$ "), so "an explosion occurs" may be taken to mean "there is an explosion," "something is an explosion," (or " $(\exists x)(\text{Explosion}(x))$ "). Events as much as continuants, after all, are individuals, even if they are quite different sorts of individuals. If you get a contradiction in supposing that an event *is* at any other time except that at which it occurs, so equally, you get a contradiction in supposing that a thing *is* at any other time except that at which it exists. And insofar as we are puzzled at how an event can have a characteristic at a time at which it isn't, we must be equally puzzled at how a thing can have a characteristic at a time at which it isn't. The fact that we need to distinguish between events and continuants does not require

that we press against events an objection which, if telling at all, is equally an objection against continuants. For continuants also acquire characteristics at times at which they don't exist. The grandfather case is a case in point.

VI

With all of this in mind, I return to the arguments against the identity theory of actions, and Boone's doing in of the bear. $at^3 - t_1$ (Boone's shooting the bear causes the death of the bear). (Now that I've introduced the $at^3 - t$ operator I won't make a point of using it. It's to be understood where appropriate.) But Boone's shooting the bear's *causing* the death of the bear is Boone's causing the death of the bear by shooting him, which in turn is Boone's killing the bear (cf. (iii), p. 224). And Boone's shooting the bear's causing the death of the bear is *not* an event over and above Boone's shooting the bear and the death of the bear. For, in general, where A_1 and E_1 are events, A_1 's causing E_1 is not itself an event (cf. (ii), p. 223). But then Boone's killing the bear is *not* an event over and above Boone's shooting the bear and the death of the bear.

So, although the truth of "Boone shoots the bear at t_1 " requires that there is an event at t_1 which is Boone's shooting the bear (supposing, as I'll do here, that Boone's shooting is not itself dissolvable into a causing), the truth of "Boone kills the bear at t_1 " does *not* require that there is an event at t_1 which is Boone's killing the bear. That is, Boone's shooting the bear occurs at t_1 , but Boone's killing the bear does not occur at t_1 , for to suppose that it does would be to suppose that Boone's shooting the bear's causing the death of the bear is itself a datable occurrent event.

The general point is that *not* every true statement such as "Boone kills the bear at t_1 " requires that there is an event which occurs at that time, t_1 . Every such true statement requires that there is an occurrent event, but it needn't be occurrent at the time indicated in the true statement. The situation is similar in the grandfather case. If Ted is the father of Fred, and Fred is the father of Ned, but Ned is born after the death of Ted, then Ted becomes a grandfather after his death. Suppose Ted dies at t_1 and Ned is born at t_2 . Then Ted is a grandfather at t_2 . But the truth of this statement

²⁰ Nicholas Rescher, "On the Logic of Chronological Propositions," *Mind*, vol. 75 (1966), pp. 75-96.

²¹ Moore, *Commonplace Book*, *op. cit.*, p. 85, p. 97.

does not require that there is at t_1 an existent individual, namely Ted, who is a grandfather of Ned. It requires that there is an existent individual, but the individual needn't exist at the time indicated in the true statement (though, of course, it could).

This is consistent with the *T.P.I.I.* For if there is an event x which is a shooting of a bear by Boone at t_1 , and there is an event y which is a killing of a bear by Boone at t_1 , there is thereby no property of x at t_1 which is not a property of y at t_1 , and there is no property of y at t_1 which is not a property of x at t_1 . And in particular, before the bear dies, it's not the case that there is an event, namely Boone's shooting the bear, which *has* the property of having occurred, and an event, namely Boone's killing the bear, which *lacks* the property of having occurred. Rather, there is an event which has the property of being a shooting of a bear by Boone, but lacks the property of being a killing of a bear by Boone. Similarly, it's not the case that Boone's shooting the bear has and Boone's killing the bear lacks, the property of having occurred at a certain time, t_1 . Rather, there is an event which has the property of being a shooting of a bear by Boone at t_1 , but which lacks the property of being a killing of a bear by Boone at t_1 . And the same considerations hold for the property of being followed n hours later by the bear's death. Since an event has the property of being a shooting of a bear by Boone at t_1 and only acquires the property of being a killing of the bear at t_1 , and since the bear dies at t_1 , the event's having the property of being a shooting is followed n hours later by the bear's death, but the event's having the property of being a killing is not. But that's not the same as saying that there is an event which occurs at t_1 , the shooting, and an event which occurs at t_1 , the killing, so that the shooting is followed n hours later by the bear's death, while the killing is not. Once again, the parallel with the grandfather case is clear. Ted is the father of Fred n years before the birth of Ned, but he is not the grandfather of Ned n years before the birth of Ned, and yet the father of Ted is the grandfather of Ned.

This, I think, takes care of the Thomson arguments against the identity of a shooting and a killing: the tense problem, the date problem, and the temporal order problem. So far forth, a shooting is a killing in a sense consistent with the *T.P.I.I.* This discussion trades on the notion that the same sorts of arguments which can be used against the identity of actions can be used against the identity

of continuants. But then, how is one to respond to Moore's charge that "this would assimilate an event to a thing which persists and has at one time a quality which it hasn't got at others"? There is still *this* way of distinguishing events from continuants: Continuants change both during and after their existence, but momentary events only change after their existence or occurrence (cf. (iii), p. 223). Another, connected, difference is that a continuant may or may not exist when it later acquires some characteristic, but a momentary event cannot occur at that time. For example, Ned's grandfather may exist at the time of Ned's birth, but Boone's shooting cannot occur (or recur) at the time of the bear's death.

Now to Goldman's arguments. Consider a difference in the effects of actions: Boone's shooting the bear causes the death of the bear at t_1 , but it is not the case that Boone's killing the bear causes the death of the bear at t_1 . So Boone's shooting has an effect which Boone's killing doesn't have. But that's precisely because, although there is an event which is a shooting of the bear by Boone at t_1 , that event just acquires the characteristic of being a killing of a bear by Boone at t_1 in virtue of its coming to cause the death of the bear at t_1 . The same holds for a case of diverse effects of a single cause. Suppose Boone's companion Simon Kenton was asleep in their camp when Boone shot the bear, and that Boone's shooting the bear caused Kenton to awake. Boone shot the bear, Boone killed the bear, and Boone awakened Kenton. But surely Boone's awakening Kenton did not cause the death of the bear, nor, though less clearly so, did Boone's killing the bear cause Kenton to awake. How then can Boone's shooting the bear be *both* Boone's awakening Kenton and Boone's killing the bear? The answer is that just as Boone's killing the bear does not cause the death of the bear, since it is in virtue of the death of the bear that Boone's shooting becomes a killing, so Boone's awakening Kenton does not cause Kenton to awake, since it is in virtue of Kenton's awaking that Boone's shooting the bear becomes an awakening. And since Kenton's awaking and the death of the bear are not causally connected, so Boone's awakening Kenton and Boone's killing the bear are not causally connected. Rather, Boone's shooting becomes his awakening Kenton when Kenton awakes, and becomes his killing the bear when the bear dies.

Yet another difference is that between causal or enabling conditions of Boone's shooting the bear and Boone's killing the bear. For example, if there

was not a sufficient charge in Boone's rifle, he may have shot the bear but not killed it. Or if he did not hit the bear in a vital spot, he may have shot without killing. Thus, a sufficient charge and hitting a vital spot are enabling conditions for Boone's killing the bear but not for his shooting the bear. But on my reading, this is just to say that these are conditions which enable Boone's shooting to become a killing, i.e., if these conditions are not satisfied, it won't be true that Boone's shooting the bear causes the death of the bear at t_1 .

Finally, I'll take up Goldman's arguments from the asymmetry and irreflexivity of "the by-relation," and from the basic-nonbasic action distinction. In the first, the idea is that if an act A of an agent S is identical with an act A' of the agent S , then if S does A' by doing A , it ought also to be true that S does A by doing A' . This is just an application of the substitutivity argument to the case of "by" contexts. The claim, then, is that, in general, if S does A' by doing A , it's not the case that S does A by doing A' ; consequently, it can't be the case that A is identical with A' . The first point to note is that "by" contexts don't take two (or more) descriptions of actions of the same kind. "Boone's killing of the bear by shooting him" won't do, but rather it must be "Boone killed the bear by shooting him." Thus there aren't two terms of the same kind to be inter-substitutable.²² I think this suffices to dispel the substitutivity argument for "by" contexts. But then, it also suggests that "the by-relation" doesn't hold between action-events at all. But perhaps it does. And if it does, then there is still Goldman's point that if there are asymmetric or irreflexive relations such as "the by-relation" between any items A and

A' , then they can't be identical. That's true enough under *P.I.I.* But not under the *T.P.I.I.* Boone killed the bear by shooting him. And it is not the case that Boone shot the bear by killing him. But this is true because Boone's shooting *became* his killing rather than conversely. The fact that an event changed in one direction rather than another is enough to account for this anomaly. While the truth of "Boone kills the bear at t_1 " requires more than just Boone's shooting the bear (or some such action on Boone's part), since it also requires the death of the bear, and while it requires that there is a killing of the bear by Boone, it *doesn't* require that there is an event which exists or occurs at t_1 over and above the death of the bear. It only requires that an event has acquired the characteristic of being a killing of the bear by Boone at t_1 . And the event which acquires that characteristic at t_1 is, as I've argued, Boone's shooting the bear. The argument from the basic-nonbasic action distinction turns on the argument from "the by-relation," since Goldman draws the distinction on the basis of "the by-relation." What this shows is that either "the by-relation" isn't a satisfactory basis for the basic-nonbasic action distinction, or that the distinction, for at least some cases, is to be understood as bearing on the direction of eventual change.

Events change by becoming causes. Shootings become killings by becoming the causes of deaths. Perhaps such changes are not *real* changes (however one characterizes a *real* change). If so, then perhaps killings are not *real* actions. But whether killings are real actions or not, they are not distinct from shootings. And what is the point.

Oberlin College

Received April 7, 1976

²² J. J. Thomson makes this point in "Individuating Actions," *The Journal of Philosophy*, vol. 68 (1971), pp. 75-776.

THE PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 27, No. 108, JULY 1977

ARTICLES

PLATO'S MIDDLE DIALOGUES AND THE INDEPENDENCE OF
PARTICULARS

F. C. White

POPPER'S THIRD WORLD

Brian Carr

ARMSTRONG AND THE INTERDEPENDENCE OF THE MENTAL

George Sher

SLAVERY, CONTENTMENT AND SOCIAL FREEDOM

G. W. Smith

DISCUSSIONS

ON MORAL NOSE

Arthur M. Wheeler

INTERESTS AND ANIMAL RIGHTS

R. G. Frey

BOOK REVIEWS

BOOKS RECEIVED

Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant, Anthony Ellis and Stephen Read.

Published for the Scots Philosophical Club by the University of St Andrews

Annual Subscriptions: Individuals £4.00 (\$10.50) Institutions £7.00 (\$17.50)

Single Issues £1.25 or \$3.00 (Institution £2.00 or \$5.00)

for orders placed directly with the Journals Manager, Scottish Academic Press,
Ltd., 33 Montgomery St., Edinburgh EH7 5JX Scotland

CLIO

An Interdisciplinary Journal of Literature, History, and the Philosophy of History is expanding for the Fall of 1977 to offer an equal concentration on philosophy of history, especially the thought of Hegel. Clark Butler is joining Robert H. Canary and Henry Kozicki as new co-editor, and J. N. Findlay, H. S. Harris, and Otto Pöggeler as new Advisory Editors. Articles and reviews are invited for these regions, as well as for CLIO's usual offerings in history-related literary criticism and literary analyses of history writing.

Subscriptions remain the same:

\$6 for individuals and \$15 for institutions.

All correspondence to

CLIO, Indiana University-Purdue University, Fort Wayne, Indiana 46805.

VI. THE MIND-BRAIN IDENTITY IMPASSE

RICHARD H. SCHLAGEL

EVER since its introduction and defence by philosophers Feigl, Place, Smart, and Armstrong, "The Mind-Brain Identity Theory" has aroused considerable philosophic clamor. At first the "Theory" tended to be rejected out of hand on the ground that it involved a simple category mistake. With few (but, as with Hobbes, important) exceptions, philosophers since Descartes had generally assumed that the inherent characteristics of conscious experiences and physiological processes were so irreducibly disparate that they necessarily required two distinct ontological settings, the mind and the body. Faced with seemingly insuperable opposition, the proponents of The Identity Theory nevertheless finally succeeded in making it both a conceptually possible and empirically plausible interpretation of the mind-body problem.

This achievement involved three stages. The *first* consisted of the conceptual analysis stage in which it was argued that while the intensional meanings of two terms or descriptive phrases might be incompatible and irreducible, this does not necessarily preclude their designating the same entity. To use a favorite example, while the terms "the morning star," "the evening star," and "the planet Venus" have different meanings because of their different intensions, we now realize that they designate the same planetary body. More generally, it has been argued that with the development of science many identities have been established between ordinary objects and processes with their macro qualities and inferred scientific entities with their radically different micro structures and properties. Thus we readily accept such expressions as "light consists of electromagnetic vibrations" (or of photons), that "heat is equivalent to the mean kinetic energy of molecules," that "water is H_2O ," that "genes are DNA and RNA molecules," that "lightning is a discharge of electrical particles," etc.

Along with the establishment of such identities, more exact scientific concepts often replace previously inadequate or approximate notions, as when the concept of gravitational force (which now has been replaced in relativity theory by the concept gravitational field) replaced the older concepts of

impetus and anima motrix, or when the concept oxygen replaced the concepts fixed air and phlogiston. It is argued analogously by proponents of The Identity Theory that the fact that mental concept terms do not have the same meaning as neurophysiological terms does not preclude the extensional identification of the two terms: e.g., that a felt pain can (at least in origin) be spatially-temporally identical with the firings of certain neurons.

Having argued for the conceptual possibility of the extensional identification of mentalistic and physicalistic terms, the *second* stage of the argument consisted in acknowledging the *de facto* empirically contingent but discoverable correlations between conscious processes and neurological processes. Just as at one time one did not know that lightning consists of electrical discharges, so for most of man's history he did not realize that conscious experiences are dependent upon chemical-electrical neuronal discharges. But the extensive research in neurophysiology, pharmacology, and psychoneurology during the 20th century has led most scientists and many philosophers to conclude that conscious processes could not occur without the normal functioning of neurological processes. In fact, the complex integrative and synthesizing activities which Kant attributed to the transcendental ego are now known to be largely dependent upon the brain.

In the *third* stage of the argument it was claimed not only that conscious experiences have been contingently correlated with neurological processes such that their referents can be identical, but that conscious experiences are "nothing but" neurological processes. According to this radical reductionistic stage, introduced in the name of scientific simplicity, it was argued that in a "complete" physicalistic or scientific account of the universe no reference to conscious processes or macro qualities would (at least in principle) occur. In this (theoretically projected) view, all linguistic expressions referring to conscious experiences such as pains, feelings, thoughts, memories, dreams, etc., as well as those referring to ordinary macro qualities such as colors and sounds, would be replaced by "successor concepts" of science designating such phy-

sical entities and processes as atomic-molecular motions, electromagnetic vibrations, changes of ionic potentials in nerve endings, patterns of neuronal firings, etc. Thus the former linguistic expressions gradually would be "replaced by" scientific terms and hence would "disappear" from our vocabulary. However radical, this position has been held (with varying degrees of explicitness or consistency) by Feigl, Place, Smart, Armstrong, Feyerabend, and Rorty.

To take the classic example, in the "Postscript" to his "Essay" on *The "Mental" And The "Physical,"* Feigl says, "I now agree with Smart (and perhaps Feyerabend) that within the conceptual frame of the theoretical natural science genuinely phenomenal (raw feel) terms have no place."¹ While this statement is made in reference to the conceptual framework of science, and thus does not necessarily preclude the retention of a phenomenal language outside of the framework of science, the proponents of The Identity Theory tend to opt for a "unitary theory" with one "monistic language." As Smart states, if one did not adopt a unitary theory, then "sensations, states of consciousness, do seem to be the one sort of thing left outside the physicalistic picture, and for various reasons I just cannot believe that this can be so."² Rorty even goes so far as to claim that like terms formerly used to designate demons, words designating sensations, and even those used to refer to common macro objects such as tables, could "in principle" be eliminated and still "leave our ability to describe and predict undiminished."³

Now for reasons which cannot be presented in this paper but which are based on the evidence provided by neurophysiological research, I accept the first two stages of the argument of The Identity Theory. Although this judgment is contingent upon whatever new discoveries or developments may occur in the future (either in the neurological sciences or in parapsychology), at least at the present the empirical evidence makes it exceedingly difficult to think of mental processes having an independent ontological status, while at the same time appearing to be for all the world dependent upon brain processes. Thus I reject as (presently) improbable an ontologically dualistic theory of the mind and the body. However, while agreeing with

the first two stages of the argument, I *do not* accept the third stage. And, in fact, it seems to me that once the full implications of this radical stage of the argument (i.e., that conscious experiences and macro qualities are "nothing but" brain processes) are made explicit, no reasonable person could accept them.

The difficulty of maintaining a consistent reductionistic position is particularly evident in the "Postscript" which Feigl added ten years later to his original "Essay." In the lengthy early "Essay" he had adopted a kind of double aspect view of consciousness and neurological processes. Asserting that conscious experiences are "features of cerebral processes," he maintained that the referents of conscious experiences are the same as those of neurological processes, though the way they come to be known are different. Thus superimposed on an ontological monism was a "double-knowledge, double-designation view." In the "Postscript," however, having been influenced in the intervening decade by such radical reductionists as Smart and Armstrong, Feigl attempted to avoid this designative and descriptive dualism, as well as the "nomological dangles" (the psychological laws needed to supplement purely physical laws), implied in such a position. As he says in the "Postscript:"

I realize that my own previous ("identity") account must also be thoroughly revised. As it stood ten years ago, it contained insuperable difficulties . . . If I had been satisfied with merely *extensional* identity, I would have been saddled with an ontology of particulars (preferably of events) with *dual properties* (i.e., phenomenal and physical). But that is hardly a step in the direction of the thoroughgoing monism I hoped to vindicate.⁴

Moreover, a thoroughly "monistic" view would eliminate the obtrusive "nomological dangles": "Inasmuch as a good and complete physicalistic . . . account of the world will contain 'successor' concepts to all phenomenal concepts, there will indeed be no 'nomological dangles' in such an account" (p. 144). Thus as Smart and Armstrong, Feigl would have liked to adopt a completely physicalistic or monistic account, although he never succeeded in doing so because he never quite accepted a com-

¹ Herbert Feigl, *The "Mental" And The "Physical"* (Minneapolis, [1958] 1967), p. 141.

² F. F. C. Smart, "Sensations and brain processes," revised version reprinted in C. V. Borst (ed.), *The Mind-Brain Identity Theory* (New York, 1970), pp. 53-54.

³ Richard Rorty, "Mind-Body Identity, Privacy and Categories," reprinted in C. V. Borst (ed.), *op. cit.*, pp. 194-195.

⁴ Feigl, *op. cit.*, p. 145. The following quotations from Feigl are also to this work.

plete "reduction" or "transformation" of phenomenal to purely physicalistic concepts.

If taken alone or out of context there are many passages which seem to indicate that Feigl *did* adopt a *completely* physicalistic or monistic account. He says, for example, "the very possibility of giving a complete physical account of the world (i.e., an account exclusively in terms of the basic concepts and laws of physics) is just that striking . . . feature of the universe and man's place in it that the advancing sciences make increasingly plausible" (p. 145). He then asks the obvious question:

... What precisely then happens to the familiar phenomenal features of the world as we know it in everyday life? And we answer—first sketchily and metaphorically: they are replaced, transformed, supplanted by the more rigorous, consistent, and explanatorily coherent and fruitful features of the world as represented by physical concepts (p. 145).

Such words as "replaced," "transformed," "supplanted," etc. would certainly imply a completely reductionistic account.

Not only will the familiar phenomenal features of the world be replaced in this "great transformation," but also the subjective experiences of time, purpose, value, free will, self, etc. As he often does, however, Feigl immediately adds that "nothing is 'explained away'—all these features are merely subjected to a redescription in a thoroughly 'detached' objective framework" (p. 155). But are they to be "redescribed," "transformed," "replaced," or what? This ambivalence is clearly evident in one passage:

My point is this: just as the commonsense (direct-realist) concepts of surface color, tone quality, flower fragrance . . . etc. are supplanted by their "successor concepts" . . . in *physical theory* . . . so the phenomenal predicates used in the description of after-images, sensations, feelings . . . etc. are *to be replaced by* the (as yet only sketchily known) neurophysiological and ultimately microphysical characterizations (pp. 141–142; italics added).

The ambivalence occurs because in the first part of the quotation Feigl asserts that macro qualities will be "supplanted" in "physical theory" by the successor concepts of physics, leaving the question open as to whether they still will have a place outside of physical theory. In the second half of the quotation, however, he omits this qualification, saying simply that phenomenal predicates are "to be replaced by" physiological descriptions.

The strongly reductionistic, physicalistic, and

monistic tendencies in Feigl are clearly revealed in a final passage which, though added at "the risk of misleading the reader by a very weak and distant analogy," yet is very indicative of one aspect of his thought. Drawing an analogy between the contrast of the egocentric and the intersubjectivistic scientific account of the world and the Ptolemaic and the Copernican view, he says:

Just as the looplike or retrograde motions of the geocentric description *disappear* once the "Copernican turn" has been accomplished, so the directly given qualities and the "privileged" egocentric terms *disappear* in the intersubjective account of the physicalistic conception (p. 158; italics added).

The term "disappear" certainly implies that in a completely physicalistic explanation of the world references to macro qualities and conscious experiences would be entirely eliminated: "They would all be 'reduced' to whatever are the concepts of the 'ultimate' physics . . ." (p. 14c).

Given the above statements one is then surprised to find that while on the one hand Feigl clearly implies that a complete explanation of the world could be given in terms of purely physicalistic concepts and laws, on the other hand he also holds that phenomenal data are essential for knowledge: "In my proposed reconstruction it is my private impressions, e.g., the shapes and colors in my visual field, which constitute 'ultimate' data of observation" (pp. 147–148). Moreover, the

data of direct experience provide the ultimate confirming or disconfirming evidence of all our factual knowledge. Purely phenomenal assertions require no other evidence than that which is "given" . . . (p. 149).

And in what appears to be a direct contradiction of the assertion quoted at the end of the previous paragraph, he says: "I still agree, of course . . . in considering clearly *intentional* . . . features as irreducible to a physicalistic description" (p. 150). Finally, in spite of what he has said previously about the "great transformation," he leaves little doubt as to the importance of the qualities disclosed in immediate experience.

While I should prefer not to irritate my tough-minded readers by waxing "metaphysical," I am tempted to say that the egocentric account, with its direct "labeling" of the qualities of experience, confronts *Being* (shudders?) as immediately as is possible in this world of ours, whereas all scientific accounts, owing to their quality-modality-invariance, deal with Being only indirectly and structurally (p. 155).

Appearances notwithstanding, my intention is not to juxtapose quotations in such a way as to embarrass Professor Feigl with apparent inconsistencies. As long as one separates the *physicalistic explanation* of natural processes from a *phenomenal description* of experience as such, there may be no explicit contradiction. Of course this would leave him with a double aspect view which he had intended to eliminate in favor of a completely physicalistic monism. But it is not surprising that Feigl was unable to reconcile these diverse tendencies in his thought. As a member of the Vienna Circle, he undoubtedly had subscribed to the phenomenalistic program of the early Positivists according to which the inferred entities of science, along with ordinary physical objects, were held to be "constructs from immediate sensory data" and consequently having no reality of their own. Now, in a 180° turn, the radical reductionists are proclaiming that the data of *immediate experience* are not ultimately real, only the inferred entities denoted by the physical framework of science. No wonder Feigl had difficulty in making a complete about-face and adopting a purely physicalistic account, as appealing as that would have been scientifically. Instead, he ambivalently maintained the *explanatory self-sufficiency* of the physicalistic framework along with the *epistemic priority* of the phenomenal framework—a clearly irreconcilable position.

In marked contrast to Feigl's ambivalence, Place, Smart, Armstrong, Rorty, and Feyerabend adopt a consistently radical physicalism. For them the elimination of human consciousness along with the phenomenal and the macro qualities of man's experience from a completely scientific account of the world is accepted with very little qualms. As Armstrong states:

What does modern science have to say about the nature of man? . . . This is the view that we can give a complete account of man *in purely physico-chemical terms* . . . We must try to work out an account of the nature of mind which is compatible with the view that man is nothing but a physical-chemical mechanism.⁵

"That everything should be explicable in terms of physics," Smart asserts, "except the occurrence of

sensations seems to me to be frankly unbelievable . . . A man is a vast arrangement of physical particles, but there are not, over and above this, sensations or states of consciousness."⁶ This confidence in a completely physicalistic account of man, in addition to being based upon the past remarkable successes of the physical sciences in explaining many physical, chemical, and biological phenomena, assumes that future developments in neurophysiology will be able to explain conscious experiences. For example, Armstrong asserts:

There are, of course, all sorts of disagreements and divergencies in the views of individual scientists. But I think it is true to say that one view is steadily gaining ground, so that it bids fair to become established scientific doctrine. This is the view that we can give a complete account of man *in purely physico-chemical terms* (p. 67).

Unfortunately for the reductionist, however, Armstrong's claim is practically without foundation. For nearly all neurophysiologists, from Sherrington to Sperry, Eccles, and Penfield, have admitted that there is absolutely *no explanation* as to how cortical patterns of neuronal discharges give rise to conscious experiences such as sensing, feeling, thinking, etc. As Karczmar has recently stated:

In fact, the more we know about the way our "brains think for us" the more difficult it becomes to relate the physical, chemical and neurophysiological processes to the very subjective process of thought as experienced by us.⁷

Or as the neurologist W. Russell Brain has asked rhetorically: "Can we imagine how the passage of electrical impulses along certain nerve fibers to an end-station of the brain can result in a sensation of pain?"⁸ Penfield summarizes the present scientific attitude as follows:

Let us consider the brain-mind relationship briefly . . . It is a boundary which, as some philosophers explain it, does not exist at all. But for the neurophysiologist there is a working boundary that does exist. Physiological methods bring him nearer and nearer to it. But he comes to an impasse, and beyond that impasse no present-day methods can take him.⁹

⁵ D. M. Armstrong, "The Nature of Mind" in C. V. Borst (ed.), *op. cit.*, p. 67.

⁶ Smart, *op. cit.*, p. 54.

⁷ A. G. Karczmar, "Introduction" in *Brain and Human Behaviour*, ed. by A. G. Karczmar and J. C. Eccles (New York, 1972), p. 2.

⁸ W. Russell Brain, *Mind, Perception and Science*, p. 66. Quoted from Jack H. Ornstein, *The Mind and the Brain* (The Hague, 1972), p. 110.

⁹ Wilder Penfield and Lamar Roberts, *Speech and Brain-Mechanisms* (Princeton, 1959), p. 8.

Given this scientific "impasse," it is instructive to see how the radical identity theorists deal with the problem. Indirectly acknowledging the lack of a scientific explanation, they attempt to solve the problem by a kind of *conceptual expurgation* and *linguistic fiat*. Consider, for example, the following statements by Smart. "Why should not sensations just be brain processes of a certain sort?" And "in so far as a sensation statement is a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes."¹⁰ Along with this extensional identification of sensations with brain processes, Smart even adopts the extreme view that the *intensional* meaning of sensations can be equated with that of neurons such that the former acquire the properties of the latter: "This does not mean that sensations do not have plenty of properties, for if they are brain processes they certainly have lots of neurological properties" (p. 61).

Thus in contrast to neurophysiologists actually doing research in the field, the radical identity theorists just adamantly maintain without any factual evidence that sensations are *nothing but* brain processes. Moreover, it is obvious in the case of Smart and Armstrong that their commitment to radical reductionism is not based on a possible scientific explanation of the mind-body problem, but on a kind of religious attitude toward Materialism or Physicalism. Smart says of the alternative to a "physicalistic picture" that "I just cannot believe that this can be so," that it "seems to me to be frankly unbelievable," and that this "is largely a confession of faith . . ." (p. 54). So without very much scientific support and without waiting for further scientific developments the radical identity theorist argues for a completely physicalistic view on the grounds that such a view is conceptually possible and more consistent with the physical sciences than any alternative position. As such, their position is just the converse of Berkeley's. While Berkeley's commitment to theism led him to deny the material world and Newtonian science in favor of an ontology based on mind and ideas, the radical identity theorist's commitment to physicalism leads him to deny consciousness and mind in favor of an ontology based on materialism.

However, in the absence of any explanation of the origin of conscious processes and macro qualities, the suggestion that all expressions re-

ferring to them could be merely expunged from our language lacks credibility. As historians of science such as Thomas Kuhn have pointed out, well established scientific theories as well as ordinary linguistic conventions are only discarded when there are available more successful explanatory concepts and theories.¹¹ Thus while one can agree with the statement of Rorty "that language changes as empirical discoveries are made," to propose in advance of any conceivable explanation of conscious processes and macro qualities their elimination, just because they fall outside the present physicalistic framework of science, seems less a scientific theory than a metaphysical dogma! Thus I find the continuation of the above statement made by Rorty to be just the opposite of what he concludes:

. . . I have constantly invoked the fact that language changes as empirical discoveries are made, in order to argue that the thesis that 'What people now call "sensations" might be discovered to be brain processes' is sensible and unconfused (p. 211).

To propose such a radical revision in our conceptual-linguistic view of human experience on the basis that "sensations might be discovered to be brain processes" is, it seems to me, neither particularly sensible nor unconfused. What does it mean, really, to suggest that sensations might be discovered to be brain processes—and nothing more? It is as if, having discovered that lightning is caused by electrical discharges and that water is composed of molecules of H₂O, one then concluded that there is no observable lightning nor drinkable water.

This in fact brings out the crucial assumption and flaw in the reductionist's argument. Most reductionists base their conclusion that conscious processes can be reduced to neurological processes on the assumption that a causal explanation having been given of lightning in terms of electrical discharges, of heat in terms of the mean kinetic energy of molecules, and of the analysis of water in terms of H₂O, these macro phenomena *can be reduced* to their microphysical causes or elements. That is, they assume that a discoverable dependence between macro phenomena and microphysical states warrants a reduction of the former to the latter. Thus Smart contends:

When I say that a sensation is a brain process or that lightning is an electric discharge, I am using 'is' in

¹⁰ Smart, *op. cit.*, pp. 55-56. The following quotations and pagination refer to the various articles in the C. V. Borst volume.

¹¹ Cf. Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 2nd edition, enlarged (Chicago, [1962] 1970), p. 77.

the sense of strict identity . . . Modern physical science tells us that lightning is a certain kind of electrical discharge . . . This, it is now believed, is what the true nature of lightning is. Note that there are not two things: a flash of lightning and an electrical discharge. There is only one thing, a flash of lightning, which is *described scientifically* as an electrical discharge to the earth from a cloud of ionised water molecules (pp. 56-57; italics added).

However, rather than providing an explanatory analogue for the mind-body problem, this macro-micro identification itself raises a number of crucial questions. In the first place, it is somewhat misleading to say, as Smart does (and as Feigl does at times), that a scientific *explanation* of observable phenomena represents an alternative *description* of the phenomena. Since the properties of an electrical discharge are quite unlike the qualities of a flash of lightning (or those of H_2O unlike that of water) it is not at all clear what is meant when it is asserted that ordinary phenomena are "described scientifically" by microphysical theories. In addition, since the intersubjective evidence for the description of a flash of lightning is radically different from the intersubjective evidence for an electrical discharge, how can it be maintained that the one "is really" the other? In short, the analogy drawn between the reductive identification of the mind with the brain and macro phenomena with microphysical states, rather than providing an explanatory analogue, really begs the whole question! We do not know anymore how (i.e., we cannot explain how) electrical discharges can appear as bolts of lightning, how the electrolysis of a mixture of hydrogen and oxygen gases can produce the properties of water, or how the motion of molecules can produce the sensation of heat, than we can explain how certain neuronal discharges can be experienced by a person as states of consciousness.

This problem becomes even more acute when, after identifying certain observable phenomena with their microphysical causes or elements, one concludes that the former do not exist at all—a necessary inference for the "disappearance" claim of the radical reductionist. Thus Rorty argues that ordinary physical objects could be eliminated from our everyday conceptual-linguistic framework, just as in the past demons have been.

If there is any point to saying that tables are nothing but clouds of molecules it is presumably to say that, in principle, we could stop making a referring use of 'table,' and of any extensionally equivalent term, and still

leave our ability to describe and predict undiminished (p. 195).

But the consequences of this claim are preposterous. If this proposal were carried through, then presumably all references to the objects, qualities, and processes of ordinary experience could be eliminated. One would refer to perceptual objects in terms of their atomic-molecular structures, to colors in terms of electromagnetic vibrations, to odors in terms of chemical-molecular properties, to sensations and pains in terms of nerve firings, etc. In short, one would act as if the ordinary world of experience no longer existed. Moreover, painting, sculpture, music, poetry, the culinary arts, etc. would also cease to exist since the various qualities, forms, and feelings of which these are a creative expression would no longer be acknowledged. To illustrate the possibility of eliminating colors, for example, Smart suggests a behavioral analysis according to which color discriminations are made on the basis solely of neurophysiological responses to physical stimuli (cf. p. 60). On Smart's view, as on Descartes' mechanistic interpretation of animal behavior, no conscious experiences or sensory qualities would mediate afferent stimulations and the resultant efferent behavioral responses.

It is because of the incredible implausibility of this view that I asserted earlier that no reasonable person could accept it once its implications were clearly set forth. And again the origin of this position lies in the assumption that because macro phenomena can be found to depend upon microphysical elements or processes, they can be reduced to them. Statements such as "tables are nothing but clouds of molecules," "man is nothing but a physical-chemical mechanism," and "man is a vast arrangement of physical particles" (assertions made respectively by Rorty, Armstrong, and Smart), have mislead their authors into making the erroneous inference that *causal explanations* constitute *reductive identifications*. But the *partial* explanation of colors by electromagnetic vibrations, of sparks by static electricity, of chemical properties by atomic structures, of sensations by nerve firings, etc., does not imply that colors, electrical sparks, chemical properties, and sensations *as experienced* do not exist. The purpose of science is to explain phenomena, not to explain them away. If every scientific explanation constituted a reduction (or disappearance) of the explained phenomena, then science would have a continuously diminishing domain. Moreover, if the natural world were made

to disappear then even science itself would be otiose, for there would be no experimental means of inquiry, as well as no phenomena to investigate and explain (it was probably the awareness of this that prevented Feigl from adopting a completely physicalistic monism).

Although the purpose of science is to explain phenomena, not to eliminate them, we should also realize that scientific explanation as we now understand it is itself limited. Scientists have been exceedingly successful in accounting for phenomena by such techniques as relating them to their antecedent causal conditions, by analyzing them into more basic elements, and by constructing conceptual models or theories from which the phenomena can be deduced (i.e., which provide a rationale for understanding the occurrence of the phenomena). But where scientists seem to be limited is in their ability to explain how or why the familiar characteristics of the world arise from these deeper contexts or levels of physical reality. Thus we can correlate conscious processes with chemical-electrical transmissions in neurons, but we cannot begin to explain how these neurological processes give rise to conscious experiences. Nor can we explain how atoms which (perhaps) consist of charged particles moving in empty space produce the macro properties of solidity, shape, texture, etc. And while we can analyze salt into atoms of sodium and chlorine and water into atoms of hydrogen and oxygen, we really do not know why the former produce the phenomenal qualities of salt and the latter those of water. Thus while the scientists in their investigations of nature keep discovering deeper levels or contexts of reality which provide a *partial* analysis and understanding of the more superficial levels, they seem unable to account for the particular qualities and structures that occur

on the higher levels. As regards the mind-body problem, the difficulty is to explain how a process of one kind (neurophysiological) can generate a phenomena of another kind (consciousness) such that while the latter is dependent upon the former it still has independent qualities and functions within the context of immediate experience.¹²

Given this limitation in knowledge it is difficult to assess the true nature of the mind-body problem. Is the problem an indication of the *inherent* limitations of scientific explanations? That is, will a new kind of understanding emerge in the future as significant in its consequences as the development of science has been during the past four centuries? Or is the problem an indication not just of an inherent limitation of *scientific* explanations, but of an irreducible, inexplicable dichotomy in reality analogous to other dichotomies in physics. Or more reassuringly but less likely, is it merely an indication of a *current* impasse in scientific explanations (as in the classic example, initial attempts in the 18th century to explain how bats avoid objects in the dark had to await the development of radar during the second world war before such an explanation was possible)? Most scientists and philosophers have assumed that the last alternative is the only reasonable one, that unanswerable questions are merely "linguistic muddles" or "pseudo-problems." But in light of the numerous current impasses in knowledge this assumption is no longer to be taken for granted. As finite creatures living in a world of seemingly limitless contexts and conditions, there is no assurance that everything is explainable. Thus it might be more honest and realistic to acknowledge this possibility than to imply that intractable problems will just disappear. Perhaps it is time to admit that final solutions of problems are a mirage: the more we strive to approach the truth, the more it recedes.

The George Washington University

Received March 26, 1976

¹² For an ingenious attempt to solve the problem from the standpoint of a "double aspect theory," see R. J. Hirst, *The Problems of Perception* (New York, 1959), ch. X.

Volume LXXXV

MIND—JULY 1977
Edited by D. W. HAMLYN

Number 339

Articles:

About Proper Names	P. ZIFF
Truth and Redundancy	D. ODEGARD
Mind-Body Problem and Indeterminacy of Translation	M. C. BRADLEY
Truth, Meaning and Decidability	N. TENNANT
Competitive Equality of Opportunity	D. A. LLOYD THOMAS
Truth and Understanding	A. MILLAR

Discussions:

Knowledge, Machines and Impoverished Beliefs	O. R. JONES
On Reading Heidegger	S. BUNGAY
Statements About Universals	F. JACKSON
Timelessness, Foreknowledge, and Free Will	D. DANIELSON
Tense and Predication	G. LLOYD
Dispositional Properties	A. D. SMITH

Book Reviews

Books Received

Notices

Published for THE MIND ASSOCIATION by BASIL BLACKWELL

Annual Subscription:

Institutions—£7.00 (inland and overseas), \$14.00 (USA and Canada)
Individuals—£4.00 (inland), £6.00 (overseas), \$12.00 (USA and Canada)
Single issues: £2.00 (inland and overseas), \$4.00 (USA and Canada)

The Monist

*An International Quarterly Journal
of General Philosophical Inquiry*

Founded 1888 by Edward C. Hegeler

Editor, Eugene FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor: Ann FREEMAN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate with the Editor in advance for Special Instructions. Papers should be 4000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

SCHEDULED PUBLICATION DATES:

Vol. 61,	No. 1	Jan.,	1978	Conceptions of the Self: East and West
Vol. 61,	No. 2	Apr.,	1978	The Philosophy of Thomas Reid
Vol. 61,	No. 3	July,	1978	Nominalism, Past and Present
Vol. 61,	No. 4	Oct.,	1978	The Philosophy and Psychology of Cognition
Vol. 62,	No. 1	Jan.,	1979	Parmenides and His Influence on Western Thought
Vol. 62,	No. 2	Apr.,	1979	Truth, Meaning, and Reference
Vol. 62,	No. 3	July,	1979	The Concept of A Person in Ethical Theory
Vol. 62,	No. 4	Oct.,	1979	Objectivity in Knowledge and Valuation

Editorial Office: Box 1908, Los Gatos, California 95030

Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, \$18.00; Individuals \$12.00.
Single copies: \$4.50. Foreign postage: Add 25 cents to single copy rate or \$1.00 to subscription rate.

VII. AM I IN THE WORLD?

ROBERT A. JAEGER

THE question whether I am in the world can be interpreted in at least two ways, since "the world" can be interpreted to mean either "the totality of material things" or "the totality of things that I do or can perceive." So the question can mean either "Am I a material thing?" or "Can I perceive myself?". These questions are quite independent. Someone could hold that I am a material thing and cannot perceive myself. On this view I am somewhat like my eyes, which are material but cannot be seen by me.¹ Alternatively, someone could hold that I am an immaterial thing and can perceive myself. This is the view that Hume attacked.

My object in this paper is to disarm what seem to me to be the most interesting arguments for the claims that I am not a material thing and that I cannot perceive myself. But first I want to say something about the four key words "I," "myself," "material," and "perceive."

(1) *I*. The token-reflexive "I" is interestingly different from the proper name "Robert Jaeger," but from this it does not follow that when I use the word "I" I am referring to something other than what "Robert Jaeger" refers to. Since I am Robert Jaeger, it follows that the two referring expressions refer to the very same thing. It is also a mistake to argue from the ineliminability of "I" to the conclusion that "I" refers to something that has a peculiar ontological status. Thomas Nagel places a great deal of emphasis on the fact that if I do not use any token-reflexive expressions then I can, in a sense, provide a complete description of the world (this will include a complete description of Robert Jaeger) but I cannot say that *I* am Robert Jaeger.² But it would be a howler to infer from this that I am not in the world. After all, if I use no token-reflexives then I cannot say that *he* is Thomas

Nagel, or that *that* is Mt. Everest, but nothing follows concerning the referents of "he" and "that."³

(2) *Myself*. "Myself" does not mean "my self," just as "itself" does not mean "its self." The latter distinction emerges vividly when one compares "This oven cleans itself" with "This oven cleans its self." So when I use the word "I" I am referring to myself, not to my self. And when I say that I can perceive myself I am not saying that I can perceive my self, or that my self can perceive itself, or that my self can perceive its self.

(3) *Material*. When I say that I am a material thing what I mean is that I have not only mental properties but also physical properties: I have a certain size, shape, mass, location, etc. Since "body" is often used as a synonym of "material thing" the point could also be expressed by saying "I am a body," but then it becomes essential to distinguish that thesis both from "I have a body" and from "I am my body," i.e., "I am the body that I have." The latter two theses are totally unclear, since it is totally unclear what "have" means here. There are three possibilities. (a) According to the dualist, "have" signifies an irreflexive and anti-symmetrical possession relation that immaterial minds bear to material bodies. On this interpretation, "I have a body" is true but "I am my body" is false. (b) The anti-dualist might say that "have" just means "am": "have" signifies the identity relation, not some possession relation. Thus "I have a body" means "I am a body" and "I am the body that I have" means "I am the body that I am," which is in turn equivalent to "I am a body." (More generally, "*Fa*" is equivalent to " $a = (ix)(Fx \ \& \ x = a)$ ".) So on this interpretation, "I have a body" turns out to be equivalent to "I am my body," and each turns out to be an un-

¹ This view is defended by Keith Gunderson in "Asymmetries and Mind-Body Perplexities" in Michael Racner and Stephen Winokur (eds.), *Minnesota Studies in the Philosophy of Science*, IV (Minneapolis, 1970), and (abridged) in D. M. Rosenthal (ed.), *Materialism and the Mind-Body Problem* (Englewood Cliffs, 1971).

² Thomas Nagel, "Physicalism" in Rosenthal, *op. cit.*, pp. 109-110.

³ Nagel suggests (p. 108) that the conviction that I am not in the world flows from "the feeling that there is a fundamental distinction between the subjective and the objective which cannot be bridged." This suggestion seems to me to be exactly right, but I think that in developing the suggestion Nagel overestimates the importance of token-reflexives.

perspicuous stylistic variant of "I am a body."
 (c) The anti-dualist might offer some third interpretation of "have," but until this interpretation is made clear it is impossible to say anything at all about either "I have a body" or "I am my body."

(4) *Perceive*. When Hume entered most intimately into himself he stumbled on various perceptions but did not observe anything other than the perceptions. Now, my claim is *not* that when I introspect I observe not only perceptions but also myself. It is thoroughly problematic to assert that when I introspect I observe perceptions, or some other kind of mental particular. It is not clear what these mental particulars are, and if there are such things it is not clear that they can be observed or perceived in any ordinary sense. The latter point is conceded by those who believe in the perception of mental particulars, for they have generally emphasized that we perceive mental particulars directly and physical particulars indirectly. But it would be less misleading to say that we do not perceive mental particulars and do perceive physical particulars. So when I say that I can perceive myself I have in mind the ordinary modes of sense perception, not introspection or some other mode of pseudo-perception.

I

One of the most interesting arguments for the thesis that I am immaterial is a very simple argument that seems to me to have been invented by Descartes. It goes as follows. "Although I think that I am a material thing, I must admit that there is a fundamental distinction between *thinking* that I am material (a *subjective* state of affairs) and *being* material (an *objective* state of affairs). Hence I could conceivably be mistaken in thinking that I am material, so (a) there is a logically possible world in which I think that I am material, but it is not the case that I am material. But thought implies existence, so (b) I exist in every world in which I think that I am material. Hence (c) there is a world in which I exist, but it is not the case that I am material. But material things are essentially material, so (d) if I am material then there is no world in which I exist but it is not the case that I am material. Hence (e) it is not the case that I am material."

I shall argue in the next section that (a) is false; in this section I shall argue that the argument is otherwise flawless.

The inference from (c) and (d) to (e) is obviously

valid, and it seems to me that the essentialist premiss (d) is true. Consider, for example, Descartes' piece of wax. There are worlds in which the wax has physical properties other than those it actually has, but no worlds in which the wax has no physical properties at all. The wax could, for example, have a different size and shape, but it must have some size and shape. Extension is, as Descartes thought, an essential attribute of the wax. There are, to be sure, important differences between pieces of wax and those material things that think, perceive, etc. But I see no route from the assumption that *x* is a material thing that has mental properties to the conclusion that *x* is just contingently material. The idea that "matter cannot think" obviously provides no such route, since that idea supports the conclusion that there are no material thinking things, not the conclusion that material thinking things are just contingently material.

The inference from (a) and (b) to (c) is also valid. We can symbolize the inference as follows:

A1. $\Diamond (I \text{ think that } I \text{ am material} \ \& \ \sim [I \text{ am material}])$

A2. $\Box (I \text{ think that } I \text{ am material} \supset I \text{ exist})$

$\therefore$ A3. $\Diamond (I \text{ exist} \ \& \ \sim [I \text{ am material}])$

Argument A exemplifies the following valid form:

B1. $\Diamond (a \text{ thinks that } P \ \& \ \sim P)$

B2. $\Box (a \text{ thinks that } P \supset a \text{ exists})$

$\therefore$ B3. $\Diamond (a \text{ exists} \ \& \ \sim P)$

To avoid an important misunderstanding I want to mention explicitly that argument A is not to be confused with argument C, which exemplifies the invalid form D:

C1. $\Diamond (I \text{ think that } I \text{ am material} \ \& \ \sim [I \text{ am material}])$

C2. $\Box (I \text{ think that } I \text{ exist} \supset I \text{ exist})$

$\therefore$ C3. $\Diamond (I \text{ exist} \ \& \ \sim [I \text{ am material}])$

D1. $\Diamond (a \text{ thinks that } P \ \& \ \sim P)$

D2. $\Box (a \text{ thinks that } Q \supset Q)$

$\therefore$ D3. $\Diamond (Q \ \& \ \sim P)$

To see that D is invalid, substitute "*a* is in pain" for '*Q*' and "*a* is in pain $\vee$ it will rain in London on June 1, 1976" for '*P*'. Then D1 is true (consider a world in which the disjunction is false but *a* thinks that it is true because he thinks that the second disjunct is true), D2 is the incorrigibility thesis, and D3 is false. If D is valid then the incorrigibility thesis is easily refuted by pointing out that its con-

junction with an obvious truth implies an obvious falsehood. It seems to me more reasonable to conclude that D is invalid than to conclude that the incorrigibility thesis can be refuted that easily.⁴

Not only does (c) follow from (a) and (b), but (b) is true. The point here is that "I think that I am material" presupposes "I exist": in those worlds in which I do not exist the former sentence is either false or neither true nor false. (It makes no difference, for our purposes, which of these verdicts is preferable.) To clarify this notion of presupposition, notice that there is a difference between saying that "*Fa*" is being used in such a way as to presuppose "*a* exists" and saying that "*Fa*" logically implies " $(\exists x)(x=a)$ " in the standard systems of quantification. In the standard systems " $(\exists x)(x=a)$ " is a valid sentence and hence is implied by every sentence, but "*a* exists" is not a necessary truth and is not presupposed by every sentence. Furthermore, the claim that "*Fa*" is being used in such a way as to presuppose "*a* exists" is compatible with the fact that "*Fa*" does not imply " $(\exists x)(x=a)$ " in free logic, a logic which is very convenient on those occasions when it seems useful to say that Hamlet walked and thought but did not exist. The presupposition theorist is free to say, for example, that if '*a*' is a fictional name then "*Fa*" does not presuppose "*a* exists." Does this not amount to the curious position that we can infer "*a* exists" from "*Fa*" only if we already know that *a* exists? No, since the assumption that '*a*' is not a fictional name does not imply that *a* exists (i.e., that '*a*' designates an existing thing). "Pegasus" is a fictional name, "Bucephalus" is not, but from this it does not follow

that there really was such a horse as Bucephalus. Very roughly, what makes "Bucephalus" non-fictional is certain facts about the beliefs and intentions of the users of the name, and from these facts nothing follows concerning the equine population of the world.⁵

II

Let us turn now to the very first step of the Cartesian argument. To place this step in proper perspective, consider Goldbach's Conjecture (GC): every even number is the sum of two primes. It seems plausible to say that there is a distinction between *thinking* that GC is true and GC's *being* true, hence it seems plausible to say that I could be mistaken in thinking that GC is true. But problems arise if we infer from this, or confuse this with saying, that there is a possible world in which I think that GC is true, but GC is not true. The problem is that on the standard view of mathematics GC is necessarily true if it is true. If we adopt this "essentialist" view of GC and suppose that there is a world in which I mistakenly think that GC is true, then we are committed to saying that GC is false.

Now we can locate the fallacy in the Cartesian argument. It seems plausible to suppose that there is a distinction between thinking that I am material and being material; it seems equally plausible to suppose that if I am material then I am so essentially. But if we make both suppositions then we must block the inference from the first supposition to the conclusion that there is a possible world in which I mistakenly think that I am material.

I want now to examine in more detail the fal-

⁴ The invalidity of D is relevant both to the philosophy of Descartes and to the mind/brain identity theory. Descartes sometimes seems to argue as follows: there is a world in which I mistakenly think that I have a body; there is no world in which I mistakenly think that I exist; hence there is a world in which I exist but it is not the case that I have a body. This is an invalid argument of form D, but a valid argument of form B may be produced by replacing the second premiss with "There is no world in which I think that I have a body, but do not exist." The two arguments are easily confused, since in each case what underlies the second premiss is the general thesis that thought implies existence.

The invalidity of argument D also means that various epistemological objections to the mind/brain identity theory are even weaker than is generally acknowledged. Suppose that I could not be mistaken in thinking that I am in pain, but could be mistaken in thinking that I am in neural state S. From this it does not even follow that I *could* be in pain without being in S. Indeed, the identity theorist might claim that it is a necessary (but synthetic and *a posteriori*) truth that my being in pain is identical with my being in S, and might then add that the identity seems contingent because we have been deceived by such fallacious arguments as D. This is one way of meeting the anti-identity arguments in Saul A. Kripke, "Naming and Necessity" in Gilbert Harman and Donald Davidson (eds.), *Semantics of Natural Language* (Dordrecht, 1972), pp. 334-342. (See also note 6 below.)

⁵ Jaakko Hintikka rightly points out (see his "Cogito, Ergo Sum: Inference or Performance?", *The Philosophical Review*, vol. 71 [1962], pp. 3-32) that the inference from "I think" to "I exist" cannot be defended by appealing to the standard systems of quantification (in which both " $Fa \supset (\exists x)(x=a)$ " and " $(\exists x)(x=a)$ " are valid sentences), but wrongly suggests that the only alternative is free logic (in which neither sentence is valid). Another alternative is presuppositional logic (in which the conditional is valid but its consequent is not). This alternative is also overlooked, or at least not explicitly acknowledged, in Fred Feldman's recent critique of Hintikka, "On the Performatory Interpretation of the Cogito," *The Philosophical Review*, vol. 82 (1973), pp. 345-363.

lacious inference from "There is a distinction between thinking that I am material and being material" to "There is a world in which I mistakenly think that I am material." This inference can be filled out in at least three ways, each of which involves mistakes of fairly general significance.

(1) *Analyticity*. "I think that I am material" does not mean the same thing as "I am material." Moreover, the meaning of the latter statement does not seem to be in any sense "contained in" the meaning of the former statement. Hence the statement "I think that I am material, but it is not the case that I am material" is not self-contradictory, but merely an example of Moore's Paradox. But if the statement is not self-contradictory, does it not follow that there is a possible world in which the statement is true? No, since "self-contradictory" is ambiguous between "analytically false" ("false in virtue of meaning") and "necessarily false" ("false in every possible world"). Goldbach's Conjecture is not analytically false: those who affirm GC do not thereby invite the suspicion that they have not fully grasped all the relevant concepts. But from this it does not follow that GC is not necessarily false, since that conclusion leads, *via* the standard view of mathematics, to the conclusion that GC is true. Similarly, we may cheerfully concede that "I mistakenly think that I am material" is not analytically false; the question remains whether that statement is not necessarily false (i.e., possibly true).

(2) *Evidence*. The belief in a fundamental distinction between thinking that I am material and being material can also be given an epistemological

twist, as in this argument: "My belief that I am material is based on sensory evidence; but the evidence does not provide *logically conclusive* support for the hypothesis that I am material; hence there is a world in which I believe, on the basis of that evidence, that I am material, but it is not the case that I am material."

The first problem with this argument is that it is very misleading to say, without further ado, that my belief that I am material is based on sensory evidence. I shall not pursue this point, since the argument also embodies a mistake about conclusive support that is of somewhat more general significance. If 'E' provides logically conclusive evidence for 'H' then it follows that there is no world in which 'E' is true and 'H' is false (i.e., 'E' strictly implies 'H'); but the converse of this is easily seen to be false. We have excellent inductive evidence for concluding that Goldbach's Conjecture is true (every "observed" even number has turned out to be the sum of two primes), but the evidence is not logically conclusive. From this we cannot infer that there is a world in which we mistakenly believe, on the basis of that evidence, that GC is true, since if there is such a world then it follows (given the standard view of mathematics) that GC is false. Similarly, we may concede that my evidence for believing that I am material is not logically conclusive; the question remains whether there is a world in which I believe, on the basis of that evidence, that I am material, but it is not the case that I am material.⁶

(3) *Causation*. If we assume that sensory evidence

⁶ The distinction between strict implication and conclusive support (which is pursued further in my "Implication and Evidence," *The Journal of Philosophy*, vol. 72 [1975], pp. 475-485) bears on some of Kripke's views in "Naming and Necessity." Suppose that Alpha is a wooden table and that I believe, on the basis of nonconclusive sensory evidence 'E', that Alpha is wooden. Suppose further, with Kripke, that since Alpha is wooden there is no world in which Alpha is not wooden. Kripke would apparently explain (see pp. 331-33) the nonconclusiveness of 'E' by saying that there is a world in which 'E' is true and there is a table (not Alpha, but a "counterpart" of Alpha) that closely resembles Alpha but is not wooden. In such a world "Alpha is wooden" is not false, but some "contingent counterpart" of that statement is false. What I wish to emphasize is that this explanation in terms of contingent counterparts is theoretically superfluous. After all, to explain the nonconclusiveness of the inductive evidence for GC we need not produce a contingent counterpart of GC.

This point undermines Kripke's attack against the suggestion that it is a necessary (but synthetic and *a posteriori*) truth that pain is identical with C-fiber stimulation. Kripke asks the identity theorist to explain why the identity statement seems contingent, to which one simple answer (but see also note 4 above) is: our evidence for believing that pain is identical with C-fiber stimulation is not logically conclusive, and from this we have fallaciously inferred that there is a world in which that evidence obtains but the identity does not. One defect of Kripke's argument is that it presupposes something like a sense-datum theory of illusion. According to the sense-datum theorist, if the straight stick looks bent then there is something that really is bent; and according to Kripke, if a necessary statement seems contingent then there is some statement—the contingent counterpart of the necessary statement—that really is contingent. The illusion of contingency stems from failing to distinguish the necessary statement from its contingent counterpart, just as one might fail to distinguish a straight stick from a bent sense-datum. A second defect of Kripke's argument is that it is hard to see the relevance of his claim (p. 339) that "to be in the same epistemic situation that would obtain if one had a pain *is* to have a pain." This amounts to saying that I cannot have nonconclusive evidence for believing that I am in pain, which is entirely compatible with the claim that I have nonconclusive evidence for believing that my being in pain is identical with my C-fibers' being stimulated.

'E' supports the hypothesis that I am material, and if we make some rather natural assumptions concerning the relation between evidence and explanation, then we are led to conclude that my being material would explain why (or would be part of an explanation why) 'E' is true. But Hume and others have insisted that all causal/explanatory statements are contingent: if 'E' is true because 'H' is true, then this is just contingently so. So if we assume that 'E' is true because I am material, we seem forced to conclude that there is a world in which 'E' is true but it is not the case that I am material.

This argument contains a fallacy rather similar to the just-discussed fallacy about evidence. If it is a contingent fact that 'E' is true because 'H' is true, then there is a world in which it is not the case that 'E' is true because 'H' is true. But given that Beta is a world in which it is not the case that 'E' is true because 'H' is true, it does not follow that Beta is a world in which 'E' is true and 'H' is false. (The converse, however, holds: if Beta is a world in which 'H' is false, then Beta is a world in which it is not the case that 'E' is true because 'H' is true. *Ex nihilo nihil fit.*) To illustrate the point, suppose that I have a stuff-meter (let us not worry about how it works) which I use to determine whether things are made of wood, ice, iron, glass, etc. I hold the meter over Alpha, the wooden table, and the needle points to "Wood" because Alpha is wooden. The Humean thesis is that there is a world in which the needle points to "Wood," but not because Alpha is wooden. (For example, perhaps there is a world in which the needle points to "Wood" because the meter is broken and the needle is stuck.) But from this thesis it does not follow that there is a world in which the needle points to "Wood" and Alpha is not wooden. Similarly, suppose that there is some sort of causal/explanatory connection between my thinking that I am material and my being material. The Humean thesis is that there is a world in which I think that I am material, but not because I am material. (For example, perhaps there is a world in which I think that I am material because the Evil Genius has implanted that thought within me.) From this thesis it does not follow that there is a world in which I think that I am material, but it is not the case that I am material.

III

In dealing with the idea that I cannot perceive myself it is essential to hang onto whatever remains

of one's robust sense of reality. After all, I can see my arms, legs, and torso, and sometimes I get a rather blurred view of the tip of my nose. I cannot see my eyes or the back of my head, but I can touch them, and can touch those parts of myself that I can see. I cannot touch my left hand with my left hand, but I can see my left hand and can touch it with my right hand. Also I can taste, smell, and hear various parts of myself. And without using any of the "external senses" I can feel that my head is tilted, that my heart is pounding, etc.

Someone might say that the above facts show that I can perceive my body, but not that I can perceive myself. The person who says this is committed to holding that I am not identical with my body, so he must either hold that I am an immaterial thing or hold that I am a material thing other than my body. In the former case the claim that I can perceive my body but not myself becomes the claim that I am an immaterial subject of perception that possesses a material body and perceives that body, but cannot perceive itself. This full-blown dualism is rather hard to take seriously. If, on the other hand, I am a material thing other than my body, then the problem is to produce a reason for saying that I cannot perceive myself that is not also a reason for saying that I cannot perceive my body. I know of no solution to this problem.

In discussions of this topic much is made of the fact that I cannot see my eyes. Less attention is paid to the question whether I can touch my left hand with my left hand, taste my tongue, smell my nose, or hear my ears. But from the fact that I cannot see *my eyes* it does not follow that I cannot see *myself*. My eyes are not me but merely a part of me, and to see a thing it is not necessary to see all its parts. And even if I could not *see* myself that would not show that I could not *perceive* myself: vision is not the only sense. Furthermore, we should not be deceived by the analogical argument "Just as my eyes cannot see themselves, so I cannot see (or perceive in some other way) myself." The first problem is that the premiss of the argument is ill-formed: my eyes do not see things, I see things with my eyes. What underlies the premiss is that my eyes cannot be in front of my eyes, so light cannot be reflected from my eyes to my eyes. But now the second problem emerges, which is that it does not follow that I cannot see (or otherwise perceive) myself. Consider this analogy. A gun cannot be in front of itself, so a bullet cannot be fired from a given gun at the very same gun. Thus one might say that a gun cannot shoot itself. But guns do not shoot things, people

shoot things with guns; and sometimes people shoot (or otherwise kill) themselves.

To say that I can perceive myself is not to say that my position with respect to myself is exactly the same as my position with respect to other things. I could look for a needle, but not for myself, in a haystack; I could come upon a friend, but not myself, in a crowded room. But these facts do not show that I cannot perceive myself. After all, I now see my left hand but I neither looked for it nor came upon it. The fact that I cannot walk around myself, though interesting, is also irrelevant. I cannot walk around my ankles but can certainly perceive them.

Other differences between my position with respect to myself and my position with respect to other things are connected with what Ryle called the systematic elusiveness of 'I'.⁷ If I attempt to write down everything that I did today then I seem to encounter an infinite regress: if I did A then I must record that I did A, and so I must record that I recorded that I did A, etc. *ad infinitum*. I encounter no such difficulty when I attempt to write down everything that you did today, hence it seems that I can provide a complete description of you but not of me. But this systematic elusiveness must be taken with at least two grains of salt.

The first point is that the concept of a complete description is notoriously problematic. The above argument strongly suggests that for me to provide a complete description it is not only necessary but also sufficient that I write down infinitely many sentences, each of which is generated from its predecessor by adding the prefix "I recorded that." The idea seems to be that if I could, *per impossibile*, complete the infinite series, then I could provide a complete description of myself. But in the exalted sense of "complete description," to provide a complete description it is not sufficient to complete the infinite series; and in the more homely sense of "complete description," it is not necessary. As for the first point, someone might say that "I recorded that I did A" is not a complete description of my recording that I did A, since nothing is said about the color of the ink, the shapes and sizes of the letters, etc. Such questions would remain unanswered even if I were to complete the infinite

series in question, so even the completed series of sentences could be considered to be an incomplete description of what it describes. If, on the other hand, "complete description" is used in its more homely sense, a complete description of one day's activities might run as follows: "I got up, did A, B, and C, wrote in my diary, and went to bed."

The second grain of salt is that even if I cannot provide a complete description of myself, I can certainly describe myself. Similarly, even if we suppose for the sake of argument that I cannot have a "complete perception" of myself⁸ (since if I perceive that I did A then I must perceive that I perceived that I did A, etc.) it does not follow that I cannot perceive myself.

IV

Having begun this paper by emphasizing the distinction between the question whether I am material and the question whether I can perceive myself, let me conclude by mentioning that the insight that animates the Cartesian argument for immateriality can also be used to support the thesis that I cannot perceive myself.⁹

According to the standard account of perception, to say that *x* perceives *y* is to say that *y* exists (or that *y* "is there"), that *x* is in a certain mental state (*x* has certain sensations, beliefs, etc.), and that *x*'s being in that state is causally connected in some appropriate way with *y*'s being there. The fact that *x* perceives *y* is thus analyzable into an objective component, a subjective component, and a causal component. The feeling that there is a fundamental and unbridgeable subjective/objective distinction leads one to say that the first two components are logically independent: no statement about *x*'s subjective state implies or is implied by any statement about the objective world. Suppose now that *x* perceives *y*. Then *y* exists and *x* is in some mental state *S*. But these consequences are logically independent, hence there is a world in which *x* is in *S* but *y* does not exist. And now the Cartesian argument applies: *x* exists in every world in which *x* is in *S*, hence there is a world in which *x* exists and *y* does not. So if *x* perceives *y* then there is a world in which *x* exists and *y* does not. To put

⁷ Gilbert Ryle, *The Concept of Mind* (New York, 1962), pp. 191-198. See also Gunderson's discussion of the self-scanning scanner: Radner and Winokur, *op. cit.*, pp. 292-300, and Rosenthal, *op. cit.*, pp. 121-123.

⁸ Compare Gunderson's claim (Radner and Winokur, *op. cit.*, pp. 298-299) that one cannot be "totally conscious" of one's self.

⁹ The point is relevant to the interpretation of *Tractatus Logico-Philosophicus* 5.634, where Wittgenstein remarks that my not being in the world is connected with the fact that "whatever we see could be other than it is."

the point more metaphysically, a subject of perception is ontologically independent of every object of its perception.

If we now suppose that x perceives x then we reach the absurdity that there is a world in which x exists and x does not exist. So a defender of the

just-sketched account of perception must either embrace the conclusion that perception is irreflexive or reconsider his feeling that there is a fundamental and unbridgeable distinction between the subjective and the objective. It is clear which alternative is preferable.¹⁰

Yale University

Received March 8, 1976

¹⁰ Early versions of this paper were read at Yale and at Livingston College, Rutgers. I thank the participants in those discussions for helpful comments, and I thank Robert Fogelin for helpful comments on a later version.

Journal for the Theory of Social Behaviour

Edited by P. F. SECORD (Department of Urban Studies, Queens College, CUNY) and R. HARRÉ (Linacre College, Oxford). Associate Editors: P. MANICAS (Department of Philosophy, Queens College, CUNY) and C. SMITH (Department of Sociology, Queens College, CUNY).

Volume 7, No. 1

April 1977

D. W. HAMLYN: The Concept of Information in Gibson's Theory of Perception

ULRIC NEISSER: Gibson's Ecological Optics: Consequences of a Different Stimulus Description

THOMAS NATSOULAS: Consciousness: Consideration of an Inferential Hypothesis

F. E. TRAINER: A Critical Analysis of Kohlberg's Contribution to the Study of Moral Thought

S. S. KOMORITA: Negotiating from Strength and the Concept of Bargaining Strength

MICHAEL MARTIN: The Philosophical Importance of Rosenthal Effect

STEPHEN P. TURNER: Complex Organizations as Savage Tribes

PHILIPPE VAN PARIJIS: Triadic Distributions and Contrepiéd Strategies: A Contribution to a Pure Theory of Expressive Behaviour

Annual Subscription: Institutions £8.50 (inland) £10.20 (overseas)

\$20.00 (USA and Canada); Individuals £5.25 (inlands) £6.30 (overseas)

\$12.50 (USA and Canada)

Orders to: JOURNALS DEPARTMENT, BASIL BLACKWELL & MOTT LTD., 108 Cowley Road, Oxford OX4 1JF, England.

Midwest Studies In Philosophy

Edited By Peter A. French, Theodore E. Uehling and Howard K. Wettstein

A philosophical annual published by the University of Minnesota, Morris.

Volume I Feb. 1976 Studies in the History of Philosophy

Volume II Feb. 1977 Studies in the Philosophy of Language

W. V. Quine: Intensions Revisited. K. Donnellan: The Contingent *A Priori* and Rigid Designators. S. Schiffer: Naming and Knowing. D. Stampe: Toward a Causal Theory of Linguistic Representation. B. Brody: Kripke on Proper Names. P. Butchvarov: Identity. F. Dretske: Referring to Events. J. Kim: Causation, Emphasis, and Events. D. Davidson: Meaning and Metaphysics. J. Wallace: Only in the Context of a Sentence Do Words Have Any Meaning. M. Levin: Explanation and Prediction in Grammar (and Semantics). D. E. Cooper: The Deletion Argument. J. O. Urmson: Performative Utterances. D. Shwayder: A Semantics of Utterance. H-N. Castañeda: On the Philosophical Foundations of the Theory of Communication. H. Wettstein: Proper Names and Propositional Opacity. H. Hochberg: Mapping, Meaning, and Metaphysics. N. Goodman: Predicates Without Properties. W. Sellars: Hochberg on Mapping, Meaning, and Metaphysics. M. Root: Quine's Thought Experiment. B. Aune: Root on Quine. S. Kripke: Speaker's Reference and Semantic Reference. *Continuing Discussion, Vol. I.* O. A. Johnson and B. Baumrin.

Volume III Feb. 1978 Studies in Ethical Theory

Papers by G. E. M. Anscombe, K. Baier, H. Bedau, D. E. Cooper, H. Fain, A. Flew, W. Frankena, D. Gauthier, A. Gewirth, G. R. Grice, R. M. Hare, G. Harman, J. Harrison, T. H. Irwin, J. Kovesi, A. Montefiore, G. Nakhnikian, D. F. Pears, J. J. C. Smart, P. Taylor, G. J. Warnock, E. Zemach, and others.

Volume IV Feb. 1979 Studies in Metaphysics

Individuals, \$10.00; Students, \$9.00; Libraries and Institutions, \$15.00. Add \$2.00 per volume for foreign or first class postage. Volume I available at \$8.50 to individuals, \$7.50 to students, and \$12.50 to libraries and institutions.

MIDWEST STUDIES IN PHILOSOPHY

University of Minnesota, Morris; Morris, MN. 56267

VIII. SETS AS OPEN SENTENCES

MICHAEL McDERMOTT

I

RUSSELL thought that sets were logical constructions out of propositional functions—he thought that sentences ostensibly referring to sets could be translated into sentences referring to propositional functions instead. By a propositional function he meant what we would call an open sentence—"a statement containing a variable x and such that it becomes a proposition when x is given any fixed determined meaning."¹ Unfortunately, this account is incompatible with Russell's detailed theory of propositional functions. His Axiom of Reducibility implies that there are indenumerably many propositional functions, whereas there are only denumerably many open sentences. If propositional functions are to be taken as anything other than sets, Quine has shown, they must be taken as attributes—entities no less abstract than sets, no less non-linguistic, and much less clear in their identity conditions.

More recently, C. S. Chihara² has pursued Russell's project of ontological reduction, but with a different theory of sets. He has shown that Wang's system Σ_ω may be interpreted as referring to open sentences. Unfortunately, Σ_ω is not widely accepted as being a satisfactory theory of sets. Also the Vicious Circle Principle, on which both Russell and Chihara base their theories of open sentences, has received no formulation in which it is both clear and plausible.

The purpose of this paper is to present an alternative method of reducing sets to open sentences. I offer first a formal, axiomatized theory of open sentences. It avoids the semantic paradoxes not by invoking any theory of types, any Vicious Circle Principle, or any notion of constructivity or predicativity. It takes, rather, a Tarskian "levels of language" approach. But in detail it is different, and superior, to the usual developments of this approach. Then a theory of sets is presented. In form it is similar to the Simple Theory of Types; but in deductive power it is weaker, in a desirable way—although

it permits the Simple Theory of Types' development of foundations for classical mathematics, it prevents all proofs of indenumerability. As suggested by its adequacy to classical mathematics, it is significantly more powerful than constructive or predicative set theories. Finally, the theory of sets is shown to be reducible, in a particularly simple manner, to the theory of open sentences.

My support for the project of reducing sets to open sentences does not derive from the nominalistic scruples which motivated Russell and Chihara. The open sentences we need are forms of inscription rather than concrete inscriptions, and there is no way of reducing the former to the latter which does not involve disreputable modal notions.

I am impressed, rather, by the need for a philosophically coherent treatment of the paradoxes of sets and those of open sentences. The currently prevailing idea that the logical and semantic paradoxes should be dealt with in quite separate and independent ways is, I think, at best too pessimistic. For there are some strong signs that a more coherent approach should work.

Many philosophers have commented on the "striking analogy" between, for example, Russell's paradox and Grelling's. It is worthwhile to remind ourselves how strong the analogy is, for since Ramsey it has been customary to play it down, as being little more than a coincidence. I wish to suggest, on the contrary, that a study of the semantic paradoxes may point to a correct treatment of the set-theoretic ones.

Intuitively, it would seem that the following schema is valid, where " $x \in y$ " is read " x is a member of y ":

$$(1) (\exists y)(\forall x)(x \in y \equiv \phi)$$

(Here ϕ is schematic for an open sentence with x as free variable.) But the schema has a logically false instance, namely

$$(2) (\exists y)(\forall x)(x \in y \equiv x \notin x)$$

This is Russell's paradox.

¹ A. N. Whitehead & Bertrand Russell, *Principia Mathematica* (Cambridge, 1927), p. 38.

² *Ontology and the Vicious-Circle Principle* (Ithaca and London, 1973).

Intuitively, it would seem that the schema (1) is valid also when " $x \in y$ " is read " y is true of x " (or, equivalently, " x satisfies y "). Truth of is a relation to objects borne by open sentences: for instance the open sentence " x is red" is true of just those objects x such that x is red. Each instance of (1) says, on this reading, that there is something true of just those objects x such that ϕ ; and surely there is—the open sentence ϕ we have just written. But (1) still has the logically false instance (2). Now, (2) says that there is an open sentence true of just those things not true of themselves. That such an open sentence would be both true of itself and not true of itself is Grelling's paradox.

Even considering this analogy by itself, it would be hard to believe that there should not be some analogy between the correct treatments of the two paradoxes. But of course we have also a strong intuitive conviction that there is an intimate *conceptual* connection between sets and open sentences. This so-called "naive" view, according to which sets are simply the extensions of open sentences, is often supposed to be discredited by Russell's paradox itself; but attention to our analogy will show that this is much too quick.

The "naive" view says, in part, that every open sentence determines a set—i.e. for every open sentence there is a set whose members are just the things satisfying the open sentence. The argument I object to simply takes schema (1) as the proper expression of this claim, and concludes that the invalidity of (1) refutes the "naive" view. But (1) cannot be the proper expression of "every open sentence determines a set": if it were, the reading of " $x \in y$ " as " y is true of x " would make (1) the proper expression of the tautology "every open sentence determines an open sentence."

Let us for a moment write " y is true of x " as " $x \delta y$." The "every open sentence determines a set" may be fairly expressed as

$$(3) (\forall z)(\exists y)(\forall x)(x \in y \equiv x \delta z)$$

When combined with the following schema of open sentence existence

$$(4) (\exists y)(\forall x)(x \delta y \equiv \phi)$$

this yields the schema (1) of set existence which Russell's paradox invalidates. But is the fault with (3) or (4)? Well, (4) itself is invalid, as shown by Grelling's paradox, so there is no need to reject (3).

So Russell's paradox does not force us to aban-

don the "naive" view of sets. If we hold the "naive" view, Russell's paradox tells us, rather, to look to our theory of open sentences and the semantic paradoxes. For if sets are the extensions of open sentences, how can we expect to find out what sets exist without first finding out what open sentences exist?

These remarks, however, are intended only to motivate the reduction project, not to advance it. I have tried to show that there *should* be a satisfactory, philosophically coherent theory of open sentences and sets; to show that there *is* such a theory is the task of the rest of the paper. My treatment of the semantic paradoxes is outlined informally in § II. In § III the formal theory of open sentences is constructed. In § IV the set theory is introduced; this section is completely independent of earlier ones. The two theories are then brought together in § V.

II

The basis of Tarskian approaches to the semantic paradoxes is the idea that semantic predicates do not belong to the languages they refer to. "True-in- L ," for instance, does not belong to L : a sentence of the form " x is true-in- L " is not a sentence of L , and hence cannot be true-in- L .

If we are to be able, in a colloquial language, to apply semantic predicates to its own sentences, we must first "split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalised language stands to its meta-language."³

Reforming English in this way, we first specify a sub-language of least extent, containing no semantic predicates: call it E_0 . Corresponding to the semantic predicates of unreformed English, we introduce semantic predicates referring to E_0 , such as "true-in- E_0 ." The addition of these predicates to E_0 yields a new sub-language E_1 . Then we introduce "true-in- E_1 " etc., whose addition yields E_2 . This process is continued indefinitely.

This is about as far as Tarski takes us, in this direction. The usual next move is to weigh in with wholesale attributions of ill-formedness of meaningfulness. For instance, " x is true-in- E_0 " would be deemed meaningless if x is not a sentence of E_0 .

This move stops the paradoxes, certainly. Suppose I write at the top of the blackboard "The

³ Alfred Tarski, "The Concept of Truth in Formalised Languages" in *Logic, Semantics, Mathematics* (Oxford, 1956), p. 267.

sentence at the top of the blackboard is not true-in- E_0 ." A little empirical investigation shows that the sentence referred to is not a sentence of E_0 , so what I have written is deemed meaningless, and implies no contradiction.

But it is an unnecessarily drastic move. Truth values may be consistently assigned to such sentences. Our sentence at the top of the blackboard, for instance, may be deemed true. Of course it won't be true-in- E_0 , since it is not a sentence of E_0 . It will be true-in- E_1 . We don't need all those attributions of meaningless to avoid the paradoxes—the hierarchy of language levels is enough. To say that a sentence is both true and not true is unsatisfactory, but there is no paradox in saying that it is both true-in- E_1 and not true-in- E_0 .

Another example: I write at the top of the blackboard, "The sentence at the top of the blackboard is false-in- E_0 ." A sentence is false-in- E_0 iff its denial is true-in- E_0 , so this sentence can't be true (i.e. it can't be true-in- E_0 or true-in- E_1 or . . .). So we must say that the sentence at the top of the blackboard is false. But it is false-in- E_1 , not false-in- E_0 .

There is no departure here from the principle of Excluded Middle. Every sentence of E_0 , for instance, is either true-in- E_0 or false-in- E_0 . And for reformed English as a whole, the schema " p or not- p " is still valid. Other key logical principles for reformed English will be⁴

x is true-in- $E_n \rightarrow p$ (substitute for x any individual name of the sentence substituted for p)

x is true-in- $E_n \rightarrow x$ is true-in- E_{n+1}

These are both one-way entailments only.

My alternative has the usual Quinean virtues of filling truth-value gaps. (Quine, strangely, accepts what I have called the usual treatment of the semantic paradoxes, in "The Ways of Paradox."⁵) And it has one special advantage over the usual treatment: although we still have to do empirical investigation to see whether "The sentence at the top of the blackboard is true-in- E_0 " is true, we no longer have to do empirical investigation to see whether it is meaningful.

The preceding treatment of "true" applies also to "true of," the key notion in the Grelling. In place of "true of" we have "true-in- E_0 of," "true-in- E_1 of" etc.; and we deem " x is true-in- E_n of y "

false if x is not an open sentence of E_n . This clears up the Grelling: "not true-in- E_0 of itself" is not true-in- E_0 of itself, but it is true-in- E_1 of itself.

One further point requires discussion in respect of "true of." We have ruled that " x is true-in- E_n of y " is false if x contains a semantic predicate referring to a language-level as high as E_n . But what about y ? Can an open sentence of E_n be true-in- E_n of an expression containing a semantic predicate referring to a language-level as high as E_n ?

I see no reason why not. The paradoxes are all avoided without this extra restriction. Furthermore, it is widely agreed that *syntactic* predicates need not be replaced by hierarchies of corresponding predicates in the same way as semantic ones. Take, then, a sentence like "The expression consisting of concatenated occurrences of the 3rd, 8th, [. . .] letters of the alphabet contains an occurrence of the 3rd letter of the alphabet." This appears to be a sentence of E_0 , irrespective of what language level the specified expression belongs to. So, in the absence of any reason for *not* doing so, we may assume that "contains an occurrence of the 3rd letter of the alphabet" is true-in- E_0 of whatever it is true of (and hence also true-in- E_1 of those things, and so on). This clears the way also for open sentences which are true-in- E_0 of everything, even themselves: for example, "either does or does not contain an occurrence of the 3rd letter of the alphabet." I repeat, such tolerance engenders no contradictions.

III

Formally, our theory of open sentences will be expressed in the categories of ordinary quantificational logic. We need just one sort of variable: each variable ranges over the entire universe of the theory—there are no types. We need an infinite list of primitive predicates— ϵ_0 , ϵ_1 , ϵ_2 etc. The atomic sentences of the theory are all formulas consisting of a predicate letter with a variable on each side. More complex sentences are constructed by the use of the usual truth-functional connectives and quantifiers.

Our formalized theory will be called *SF*, for "semantic foundations," since it is a semantic theory which provides foundations for set theory, and hence for mathematics.

The intended interpretation of *SF* consists in

⁴ The subscript n here, and later, is schematic, not a bindable variable: " $(\exists n)(x$ is true-in- E_n)," for instance, is not meaningful in reformed English.

⁵ *The Ways of Paradox and Other Essays* (New York, 1966), p. 9f.

reading " $x \epsilon_n y$ " as " x satisfies-in- $L_n y$," or " y is true-in- L_n of x ." Its universe consists of the open sentences of a certain language L . L is a formalised quantificational language containing the predicates ϵ_n , so that all sentences of SF are sentences of L . However—and this is most important— L has other predicates as well: the universe of SF includes all the open sentences of SF , and others as well.

The importance of this may be illustrated by reverting for a moment to our discussion of reformed English. Consider the language SE , which consists of all the sentences of reformed English which contain no other predicates than "true-in- E_0 of," "true-in- E_1 of," etc. One of these sentences is:

Something is true-in- E_0 of itself.

As argued above, this is a truth of reformed English: one open sentence which makes it true is "either does or does not contain an occurrence of the 3rd letter of the alphabet." There is, however, no open sentence of SE which makes it true: all open sentences of SE contain semantic predicates and hence are true-in- E_0 of nothing. It is essential for the truth of some sentences of SE that they refer to a language wider than SE .

Returning now to our language L , the wider language referred to by the sentences of SF : it, too, contains in its lowest level L_0 syntactic predicates sufficient to provide a description of each open sentence of the entire language L . For every open sentence x , there is an open sentence true-in- L_0 just of x . It is therefore true, on the intended interpretation, that if x and y are objects of which all the same things are true-in- L_0 , they are identical, and hence anything true-in- L_n of x is true-in- L_n of y :

$$(i) (\forall z)(x \epsilon_0 z \equiv y \epsilon_0 z) \supset (x \epsilon_n w \supset y \epsilon_n w)$$

Every instance of the following schema will be true on the intended interpretation:

$$(\exists y)(\forall x)(x \epsilon_n y \equiv \phi)$$

where— ϕ contains no ϵ_m , $m \geq n$; and
— x is the only free variable in ϕ .

Each such instance asserts the existence of the exhibited open sentence ϕ . For since the only predicate letters in ϕ are ϵ_m , $m < n$, it is an open sentence of L_n , and hence is true-in- L_n of any object x such that ϕ .

To accommodate general statements of open

sentence existence we may admit to ϕ free variables other than x —even after occurrences of ϵ_n , as for example in

$$(\exists y)(\forall x)(x \epsilon_n y \equiv x \epsilon_n z \ \& \ x \epsilon_n w)$$

More precisely, we accommodate such statements by strengthening the above schema to:

$$(iia) (\exists y)(\forall x)(x \epsilon_n y \equiv \phi)$$

where— y is not free in ϕ ;

— ϕ contains no ϵ_m , $m \geq n$; and

—every occurrence of ϵ_n in ϕ is followed by a free variable.

The third condition means that, for example, only the first of the following sentences is an instance of (iia):

$$(\exists y)(\forall x)(x \epsilon_0 y \equiv (x \epsilon_0 z \equiv x \epsilon_0 z))$$

$$(\exists y)(\forall x)(x \epsilon_0 y \equiv (z)(x \epsilon_0 z \equiv x \epsilon_0 z))$$

$$(\exists y)(\forall x)(x \epsilon_0 y \equiv x \neq_0 x)$$

Clearly, all (closures of) instances of (iia) will be true on the intended interpretation.

A further, and crucial, requirement for L is that, for each n , it should contain in its sub-language L_n all the predicates inductively definable in terms of predicates of L_n : each of its language levels must be closed under inductive definition.

What this means is best explained by examples. "Natural number" is inductively definable in terms of O and "successor": anything is a n.n. iff its being so follows from the stipulation that O is a n.n. and the successor of any n.n. is a n.n. Now suppose that we can (directly) define O and "successor" in L_5 . Then our requirement for L says that L_5 must also contain the predicate " x is a n.n."—i.e. it must contain an open sentence true-in- L_5 of just the natural numbers. The reason this is a substantial further requirement for L is that no open sentence of L_5 which is expressible in SF is true-in- L_5 of just the natural numbers. In SF , we have indeed the following open sentence, which is true of just the natural numbers:

$$(\forall z)(O \epsilon_5 z \ \& \ (\forall w)(w \epsilon_5 z \supset S'w \epsilon_5 z) \supset x \epsilon_5 z)$$

But this is an open sentence of L_6 , not of L_5 . Given a language L which fulfills our requirement, we can assert

$$(5) (\exists y)(\forall x)(x \epsilon_5 y \equiv (\forall z)(O \epsilon_5 z \ \& \ (\forall w)(w \epsilon_5 z \supset S'w \epsilon_5 z) \supset x \epsilon_5 z))$$

But the open sentence y here asserted to exist is not the one exhibited on the right of the " $\equiv$ ": it is

another, true-in- L_5 of the things the exhibited one is true-in- L_6 of. Our requirement introduces a further respect in which the vocabulary of L must outrun that of SF .

Another example: Let us call an open sentence a "counter" iff it is true-in- L_0 of nothing or true-in- L_0 of a counter; this is an inductive definition of "counter" in terms of the predicate "true-in- L_0 of," which is a predicate of L_1 . Then, if L is a language of the kind specified for the intended interpretation, there is an open sentence of L_1 true of just the counters:

$$(\exists y)(\forall x)(x \in_1 y \equiv x \text{ is a counter})$$

The counters are then specifiable as the open sentences which satisfy every open sentence z of L_1 which is such that (i) z is true-in- L_1 of anything true-in- L_0 of nothing, and (ii) z is true-in- L_1 of anything true-in- L_0 of something z is true-in- L_1 of. So we may assert

$$(6) (\exists y)(\forall x)(x \in_1 y \equiv (\forall z)((\forall w)((\forall v)v \notin_0 w \supset w \in_1 z) \& (\forall w)(\forall v)(w \in_1 z \& w \in_0 v \supset v \in_1 z) \supset x \in_1 z))$$

It may be shown that if L meets this further requirement all (closures of) instances of the following schema will be true:

$$(ii) (\exists y)(\forall x)(x \in_n y \equiv \phi)$$

where— y is not free in ϕ ;
— ϕ contains no ϵ_m , $m > n$; and
—if a bound variable occurs following an occurrence of ϵ_n , every other occurrence of that variable either follows an occurrence of ϵ_n , or is in a quantifier.

Thus (5) and (6) are instances of (ii), though not of (iia).

As the non-logical axiom schemas of SF we adopt (i) and (ii). For purposes of later reference, we also specify a weaker system $SF(a)$, which has (iia) in place of (ii).

The intended model for SF is a hierarchical language of the sort described, in which each level is closed under inductive definition. Models for $SF(a)$, however, need contain none of the inductively definable open sentences.

IV

My theory of sets is most conveniently introduced by comparison with the Simple Theory of Types (STT).

The variables of STT all have numerical superscripts. Its atomic sentences are all of the form $x^n \in y^{n+1}$. More complex sentences are constructed by the use of the usual truth-functional connectives and quantifiers. It has an axiom schema of extensionality

$$(\forall z^n)(z^n \in x^{n+1} \equiv z^n \in y^{n+1}) \supset (x^{n+1} \in u^{n+2} \supset y^{n+1} \in w^{n+2})$$

and an axiom schema of set existence

$$(\exists y^{n+1})(\forall x^n)(x^n \in y^{n+1} \equiv \phi)$$

where y^{n+1} is not free in ϕ .

Russell derived STT from the Ramified Theory of Types with the Axiom of Reducibility. The effect of his definitions is that talk of sets is translated into talk of predicative propositional functions which does not distinguish between coextensive ones. The same definitions can be applied to the Ramified Theory of Types without the Axiom of Reducibility, and the result is the so-called Constructive Theory of Types (CTT). It differs from STT only in having the following weaker axiom schema of set existence:⁶

$$(\exists y^{n+1})(\forall x^n)(x^n \in y^{n+1} \equiv \phi)$$

where— y^{n+1} is not free in ϕ ; and

— ϕ contains no variable with superscript $> n+1$, and no bound variable with superscript $\geq n+1$.

The Axiom of Reducibility was introduced into the Ramified Theory of Types mainly because the weaker system provided inadequate foundations for classical mathematics. The difference it makes can be seen much more simply by comparing STT with CTT .

The most notable failure of CTT is in connection with the Least Upper Bound principle for real numbers. Let "... x ..." be some open sentence formulable in the classical theory of real numbers; then it is an axiom of that theory that if there is a real no less than all reals x such that "... x ...", there is a least real no less than all reals x such that "... x ...". The trouble is that, although there is a system of definitions which makes all classical theorems about reals theorems of STT , it does not make most of these statements of the existence of least upper bounds theorems of CTT . Our system of definitions, following Dedekind, construes reals as nested sets of ratios: $\sqrt{2}$, for example, is construed as the set of all ratios whose squares are less than 2/1. Suppose this makes reals sets of type ϵ . Then every

⁶ My presentation of STT and CTT is modelled on that of Quine, *Set Theory and its Logic* (Cambridge, Mass., 1965), § 36.

sentence of the classical theory of reals is translated into a sentence of our type theory which contains no variables of type higher than k . " $x^k \leq y^k$," in particular, becomes " $(\forall z^{k-1})(z^{k-1} \in x^k \supset z^{k-1} \in y^k)$." And the least real no less than all reals x^k such that $\dots x^k \dots$ turns out to be the set of ratios y^{k-1} such that $(\exists x^k)(y^{k-1} \in x^k \ \& \ \dots x^k \dots)$. Now the sentence which says that the least upper bound exists is

$$(7) (\exists z^k)(\forall y^{k-1})(y^{k-1} \in z^k \equiv (\exists x^k)(y^{k-1} \in x^k \ \& \ \dots x^k \dots))$$

and this is an axiom of *STT*, but not of *CTT*.⁷

Another branch of mathematics, sometimes called classical, to which *CTT* is inadequate is that in which the existence of indenumerable totalities is asserted. On this point, however, I can do no better than endorse Quine's sentiments: "This obstruction of Cantor's theorem might itself be seen as a virtue. The ins and outs of infinite arithmetic have claimed attention because they seemed forced on us by the same intuitively acceptable principles that underlay classical mathematics. If our constructive set theory were adequate to all reasonable demands of really classical mathematics, we might see its inadequacy to infinite arithmetic as a happy escape."⁸

I now propose a theory of sets which combines the advantages of *STT* and *CTT*. Intermediate in power between them, it may be called *ITT*. It does not share the inadequacy of *CTT* to least bounds: (7) is an axiom. In fact the normal development of the theories of natural, rational, real and complex numbers, real functions etc. may be carried through exactly as in *STT*. But, like *CTT*, it prevents the proof of Cantor's theorem.

ITT is like *STT* and *CTT* except for having the set existence axiom schema

$$(\exists y^{n+1})(\forall x^n)(x^n \in y^{n+1} \equiv \phi)$$

where— y^{n+1} is not free in ϕ ; and

— ϕ contains no variable with superscript $> n+1$.

V

The following definitions provide an effective means for translating the sentences of *ITT* into

sentences of *SF*. Under this translation, every theorem of *ITT* becomes a theorem of *SF*, and every theorem of *CTT* becomes a theorem of *SF(a)*.

x_n^m for x_j where— x_j is the j^{th} variable of *SF*;
— x_n^m is the n^{th} variable with
superscript m in *ITT*; and
— $j = 2^n(2m+1)$.

(The point of this definition is to ensure that each variable of *ITT* is assigned a distinct variable of *SF*. Now the unsubscripted " ϵ " is defined; the definition is contextual, and depends on the superscripts of the flanking variables.)

$x^0 \epsilon y^1$ for $x^0 \epsilon_1 y^1$
 $x^1 \epsilon y^2$ for $(\exists z^1)((\forall w^0)(w^0 \epsilon x^1 \equiv w^0 \epsilon z^1) \ \& \ z^1 \epsilon_2 y^2)$
 $x^2 \epsilon y^3$ for $(\exists z^2)((\forall w^1)(w^1 \epsilon x^2 \equiv w^1 \epsilon z^2) \ \& \ z^2 \epsilon_3 y^3)$ etc.

Roughly speaking, these definitions identify each set of type n in *ITT* with an open sentence of level n in *SF*. The sets of type n in *CTT*, similarly, are identified with open sentences of level n in *SF(a)*. The extra sets, recognized in *ITT* but not in *CTT*, are identified with the inductively definable open sentences, recognised in *SF* but not in *SF(a)*.

In the development of classical mathematics in *STT*, certain key results are conditional on an "Axiom of Infinity," which says that there are infinitely many entities of lowest type. To get these results unconditionally, this assumption has to be made a genuine axiom of the system. Whitehead and Russell preferred to introduce it as an hypothesis, when needed, because of its apparently contingent character. In *ITT*, also, the Axiom of Infinity is required. But we need have no reservations about making it an extra genuine axiom of *ITT*. For the definitions given above, which make all the other theorems of *ITT* theorems of *SF*, do the same for the Axiom of Infinity. The whole system *ITT*+Ax. Infin. is definable in *SF*.

I have shown how the extra strength of *ITT* over *CTT*, which makes it adequate to mathematics, may be linked with *SF*'s assumption of the existence of inductively definable open sentences. This assumption is, I think, the chief novelty of *SF*. To conclude, I would like to suggest that a

⁷ The Least Upper Bound principle is often formulated "every bounded set of reals has a least bound." I have followed the usual practice of interpreting the apparent reference to sets as a use of what Quine calls the "virtual theory" of sets (see *Set Theory and its Logic*, § 2f.)—i.e. as a way of expressing a *schema*, whose instances refer to reals only, not sets. For if the reference to sets were genuine, the principle would need to be supplemented with a theory telling us what sets existed, before we could use it to infer the existence of any particular least bound; and then there would not be one "classical theory of reals," but many, incorporating various set theories.

⁸ Quine, *ibid.*, p. 265.

similar assumption in the Ramified Theory of Types would serve most of the purposes for which the Axiom of Reducibility was introduced, and without precluding the construal of propositional functions as open sentences. However, this would still not make it as good a theory of open sentences

as *SF*: the basic obscurity of the Vicious Circle Principle would remain, as would the cumbersome apparatus of orders and types (compared with *SF*'s elegant first-order logic), and the Axiom of Infinity would still be required as an independent, ad hoc assumption.⁹

Sydney University

Received March 17, 1976

⁹ The language *L*, referred to on the intended interpretation of *SF*, is supposed to be such that each level L_n is closed under inductive definition. But didn't Tarski show that the satisfaction predicate for a language is inductively definable in terms of the predicates of the language? And won't this mean that the Grelling paradox arises in each L_n , and hence that *SF* is inconsistent?

No. Each of the L_n , being closed under inductive definition, will contain infinitely many primitive predicates. But Tarski's method only applies to languages with finitely many. His sweeping claim, that the method works for *every* formalised language of finite order (*ibid.*, p. 265), is wrong.

But the final answer to worries about *SF*'s consistency is the following proof, suggested by Dana Scott.

Let *SF* be expanded to *SF* + *I* by the addition of an identity predicate. Then *SF* + *I* has a normal model in the integers. Proof:

- (a) In *SF* + *I*, axiom schema (i) follows from (ii). So we only need to provide an interpretation of the ϵ_n which satisfies (ii).
- (b) Consider first those instances of (ii) which contain no other predicates than = and ϵ_0 . These have a model in 2nd order arithmetic (interpret = as identity between integers, and ϵ_0 as membership). Then there is also a non-standard ω -sorted model in which the sets are denumerable. Choose some enumeration of them. Then there is also an isomorphic model in which the m^{th} set is replaced by the integer m ; this fixes the interpretation of ϵ_0 .
- (c) Now consider the new instances of (ii) which contain no ϵ_n , $n > 1$. Since ϵ_0 is an arithmetic predicate, there have a model in 2nd-order arithmetic (ϵ_1 now being interpreted as membership). Operate as above to get an arithmetic interpretation for ϵ_1 .
- (d) And so on.

BOOKS RECEIVED

- Peter Angeles (ed.), *Critiques of God* (Prometheus Books: Buffalo, 1976), 371 pp. \$13.95.
- T. J. Blakeley (ed.), *Themes in Soviet Marxist Philosophy* (D. Reidel Publishing Company: Dordrecht, 1975), 224 pp.
- Myles Brand (ed.), *The Nature of Causation* (University of Illinois Press: Urbana, 1976), 387 pp. \$10.95.
- John Bricke (ed.), *Freedom & Morality* (University of Kansas Publications: Lawrence, 1976), 210 pp.
- Philipp D'Arcy, *l'argent et le pouvoir* (Presses Universitaires de France: Paris, 1976), 161 pp.
- Richard T. de George and James P. Scanlan (eds.), *Marxism and Religion in Eastern Europe* (D. Reidel Publishing Company: Dordrecht, 1976), 181 pp.
- Carolyn Eisele (ed.), *The New Elements of Mathematics by Charles S. Peirce* (Humanities Press: Atlantic Highlands, N.J., 1976), 260 pp. \$29.50
- Robert J. Fogelin, *Wittgenstein* (Routledge & Kegan Paul: London and Boston, 1976), 223 pp. \$11.00
- Charles Frankel (ed.), *Controversies and Decisions* (Russell Sage Foundation: New York, 1976), 299 pp.
- R. Baine Harris (ed.), *The Significance of Neoplatonism* (State University of New York Press: Albany, 1976), 370 pp. \$20.00
- Heele Lambridis, *Empedocles: A Philosophical Investigation* (The University of Alabama Press: University, 1976), 160 pp. \$8.00
- William Leiss, *The Limits to Satisfaction: An Essay on the Problem of Needs and Commodities* (University of Toronto Press: Toronto and Buffalo, 1976), 159 pp. \$4.50.
- Robert McRae, *Leibniz: Perception, Apperception, and Thought* (University of Toronto Press: Toronto and Buffalo, 1976), 148 pp. \$15.00.
- Alexander Rosenberg, *Microeconomic Laws* (University of Pittsburgh Press: Pittsburgh, 1976), 236 pp.
- A. Sarlemijn, *Hegel's Dialectic* (D. Reidel Publishing Company: Dordrecht, 1975), 189 pp.
- Kenneth Sayre, *Cybernetics and the Philosophy of Mind* (Humanities Press: Atlantic Highlands, N.J., 1976), 265 pp. \$15.75
- George J. Stack, *On Kierkegaard: Philosophical Fragments* (Humanities Press: Atlantic Highlands, N.J. 1976), 127 pp. \$5.00.
- Teorema, "Aspectos de la Filosofia de W. V. Quine (Universidad de Valencia: Valencia, 1976), 168 pp.
- Tulane Studies in Philosophy, Volume XXIV (1975), 130 pp. \$6.50.
- Paul Weingartner, *Wissenschaftstheorie II, 1 Grundlagenprobleme der Logik und Mathematik* (Problemata Frommann-Holzboog: Stuttgart, 1976), 358 pp.
- Andrew Woodfield, *Teleology* (Cambridge University Press: New York, 1976), 232 pp.
- Zygmunt Ziembski, *Practical Logic* (D. Reidel Publishing Company: Dordrecht and Boston, 1976), 437 pp.

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

AN INTRODUCTION TO MODAL LOGIC

by E. J. LEMMON

In collaboration with DANA SCOTT

Edited by KRISTER SEGERBERG

Monograph No. 11

Oxford 1977

This is the latest title in this very impressive monograph series, and we shall be publishing it in March 1977. The text is a collaboration between two well-known logicians, who have produced a lucid introduction to one of the fastest growing branches of modern logic. Not only will the beginner find the book particularly helpful, but it also provides a synthesis of work in the field that will interest advanced students.

To enable you to judge for yourself the range and importance of this series, a complete list of the Monographs is given on the inside back cover. The last issue supplied to institutions on subscription was No. 7 (1973); as from 1977, the price of all monographs to institutions will be \$9.00 (£4.50 to UK subscribers, £5.40 to those overseas). We regret to announce that also from 1977 it will no longer be possible to supply the Monographs free of charge to individual subscribers, and that the cost to them will be \$4.50 (£2.25 to UK subscribers, £2.70 to those overseas).

(Continued overleaf)

(Continued from previous page)

COMPLETE LIST OF MONOGRAPH SERIES

- No. 1 (1968) **STUDIES IN MORAL PHILOSOPHY.** Essays by David Braybrooke, Lawrence L. Haworth, G. P. Henderson, Jesse Kalin, Kai Nielsen, Jerome B. Schneewind, Michael Stocker.
- No. 2 (1968) **STUDIES IN LOGICAL THEORY.** Essays by James W. Cornman, Montgomery Furth, Alan Hausman and Charles Echelbarger, Jaakko Hintikka, Robert C. Stalnaker, Ted Honderich, Colwyn Williamson.
- No. 3 (1969) **STUDIES IN THE PHILOSOPHY OF SCIENCE.** Essays by Peter Achinstein, Joseph Agassi, Simon Blackburn, Mario Bunge, Bernard R. Grunstra, Keith Lehrer, D. H. Mellor, Lawrence Sklar, Michael Slote, Stephen Spielman.
- No. 4 (1970) **STUDIES IN THE THEORY OF KNOWLEDGE.** Essays by Norman Malcolm, W. Donald Oliver, Peter Unger, John L. Pollock, John Knox, Jr., Alan R. White, Frederick Stoutland.
- No. 5 (1971) **DESCARTES' PHILOSOPHY OF NATURE.** By James Collins.
- No. 6 (1972) **STUDIES IN THE PHILOSOPHY OF MIND.** Essays by Harold Brown, Haskell Fain and A. Phillips Griffiths, O. H. Green, Donald F. Henze, Moreland Perkins, Alan R. White.
- No. 7 (1973) **STUDIES IN ETHICS.** Essays by Charles V. Blatz, Claudia Card, Norman O. Dahl, R. L. Franklin, James D. Wallace, A. D. Woozley.
- No. 8 (1974) **STUDIES IN MODALITY.** By Nicholas Rescher with the collaboration of Ruth Manor, Arnold vander Nat and Zane Parks.
- No. 9 (1975) **STUDIES IN EPISTEMOLOGY.** Essays by Dennis M. Ahern, Robert Almeder, Jonathan Bennett, Fred R. Berger, Peter A. French, Margaret Gilbert, David K. Glidden, Mark Pastin, E. E. Sleinis.
- No. 10 (1976) **NEGATION AND NON-BEING.** By Richard M. Gale.

the review of metaphysics

a philosophical quarterly

MARCH 1977 | VOL. XXX, No. 3 | ISSUE No. 119 | \$5.00

articles

SUSAN HAACK
ALFONSO GOMEZ-LOBO

Lewis' Ontological Slum
Are Hypotheses in the *Posterior Analytics* Assumptions of
Existence?

JAMES COLLINS
DANIEL GUERRIÈRE
ANNETTE C. BAIER
DAVID L. NORTON

Kant's *Logic* as a Critical Aid
With What Does Hegelian Science Begin?
The Intentionality of Intentions
Can Fanaticism Be Distinguished from Moral Idealism?

critical studies

SETH BENARDETE
HENRY ROSEMONT, JR.

The Grammar of Being
Gathering Evidence for Linguistic Innateness

books received

DAVID C. BLAKE
and Staff

Summaries and Comments

philosophical abstracts

announcements

Individual Subscriptions \$12.00

Institutional Subscription \$18.00

Student Subscriptions \$7.00

Jude P. Dougherty, Editor

The Catholic University of America, Washington, D. C. 20064
Twenty-Year Cumulative Index, 1947-1967 is available for \$3.00 a copy.

RATIO

edited by STEPHAN KÖRNER

Volume XVIII

December 1976

No. 2

Kant on Existence

J. MICHAEL YOUNG

How to Dispense with Things in Themselves (II)

M. S. GRAM

Freedom of Indifference

NICHOLAS NATHAN

The Source of the River of Time

FERREL CHRISTENSEN

On the Alleged Importance of Moral Rights

WILLIAM NELSON

Mentezin: On the Problem of the Rehabilitation
of Criminals

KEVIN MCCORMICK

Reviews

Annual Subscription: £8.00 (inland) £9.00 (overseas) \$24.00 (U.S.A. and Canada)

Published by BASIL BLACKWELL, Oxford

Orders to: Journals Department, Basil Blackwell & Mott Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

INQUIRY

Editor: Alastair Hannay
Founding and Consulting Editor: Arne Naess

Volume 19 (1976), No. 1 (Spring)

Dualism and Bodily Movements
What is Wrong with Incest?
Spiral Dependence between Theories and Taxonomy
Reductionism and Structural Anthropology
The Training of Socrates
Marx's Analysis of Commodity Exchange: A Reply to Carver
The Other Side of Ecology in Ancient Greece: Comments on Hughes
E. Gellner: *Legitimation of Belief*

Thomas W. Bestor
Jerome Neu
Berent Enç
Ivan Strenski
Philip B. Wright
Ulrich Steinvorth
John Rodman
Peter Mew

Volume 19 (1976), No. 2 (Summer)

The Foundations of Cognitive Activity
Justice and the Family
Rawls and Gandhi on Civil Disobedience
Toward a Theory of Moral Debt. Pt. 2. The Factual Grounds.
A: The Good and Human Freedom
The Durability of Impossible Objects: A Reply to Lambert
On 'The Durability of Impossible Objects'
Charles Taylor: *Hegel*
Steven Lukes: *Power: A Radical View*

Knut Erik Tranøy
Francis Schrag
Vinit Haksar
Morris B. Storer
Richard Routley
Karel Lambert
G. H. R. Parkinson
Akeel Bilgrami

Subscription: £6.25, \$15.00 (institutions): £5.00, \$12.00 (individuals)

Published quarterly by UNIVERSITETSFORLAGET, P.O. Box 7508,
Skillebekk, Oslo 2, Norway

DIALOGUE

Canadian Philosophical Review
—Revue canadienne de philosophie

Vol. XVI — No. 2

Juin/June 1977

Rédaction/Editors: François Duchesneau & John Woods

ARTICLES

- Beyond Materialism and Back Again W. E. COOPER
The Surprise Examination Paradox MICHAEL STACK
La position des entités théoriques GILLES LANE
La conscience de l'observateur: de la physique théorique à la logique
mathématique YVON PROVENÇAL
Indétermination de la traduction et indétermination de la référence:
quelques remarques sur les quatre premiers essais du recueil *Ontological*
Relativity JACQUES LARGEAULT

NOTE — DISCUSSION

- Immorality and Self-Deception: A Reply to Béla Szabados M. W. MARTIN

CRITICAL NOTICES — ETUDES CRITIQUES

- Remarques sur le Rousseau de Victor Goldschmidt GUY LAFRANCE
Anarchy, State and Utopia, by Robert Nozick JAN NARVESON
The Structure of Scientific Theories, edited with a critical introduction by
Frederick Suppe STEPHEN F. SAVITT

AUSLEGUNG

A Graduate Journal of
PHILOSOPHY

Editors:

Debra Nails
Joe D. VanZandt

Board of Advisory Editors: Richard Cole, Richard T. DeGeorge,
Nancy D. Simco, Anthony C. Genova, Rex Martin

AUSLEGUNG is a new journal published three times a year at the University of Kansas. The journal is intended as a forum for the expression of any and all philosophical perspectives. The editors are primarily interested in publishing the work of new Ph.D.'s and students pursuing the Ph.D. degree, but all papers of philosophical interest will be considered.

Annual Subscription Rates: \$6.00/individuals, \$4.00/students, \$10.00/institutions.

AUSLEGUNG is happy to acknowledge the financial support which it receives from the Student Senate and the Graduate Student Council of the University of Kansas.

Manuscripts, inquiries and subscriptions should be addressed to:

AUSLEGUNG
Department of Philosophy
University of Kansas
Lawrence, Kansas 66045



Philosophy

The Journal of the Royal Institute of Philosophy

Published by Cambridge University Press

Edited by Renford Bambrough

Vol. 52 No. 199

Appearance and Reality

Consciousness and Life

Mind, Brain and Mental Illness

Words and Intentions

Philosophers Discuss Education

DISCUSSION

NEW BOOKS

January 1977

John Wisdom

Gareth B. Matthews

Leslie Stevenson

Edmond Wright

R. F. Holland

BOOKNOTES

Volume 52 (1977) £14.00 (\$32.50 U.S.A. and Canada). Single parts £4.00 (\$9.50 U.S.A. and Canada) each.

Enquiries about advertising in the Journal should be sent direct to the Press.

CAMBRIDGE UNIVERSITY PRESS

BENTLEY HOUSE, 200 EUSTON ROAD, LONDON NW1 2DB

32 EAST 57th STREET, NEW YORK, N.Y. 10022

The Journal of Philosophy

Subscriptions at \$10.00/year; \$7.00 to students
\$16.00 to libraries and institutions

The trustees and editors of the *Journal of Philosophy* announce a change in scheduling: as of January 1977, the *Journal* will be published monthly. There will be no change in the number of pages of copy published, since the monthly issues will be equivalent to the earlier double issues, and special issues will be published when needed. No change in price is anticipated.

Complete back volumes and all separate issues of the *Journal* are available, back to its founding January 7, 1904. Also, a cumulative *Fifty-year Index, 1904-1953* and *Ten-year Supplement 1954-1963*.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

GRAZER PHILOSOPHISCHE STUDIEN INTERNATIONALE ZEITSCHRIFT FÜR ANALYTISCHE PHILOSOPHIE

gemeinsam mit

K. ACHAM, R. M. CHISHOLM, R. FREUNDLICH, E. TOPITSCH und O. WEINBERGER

herausgegeben von Rudolf HALLER

Board of Consulting Editors: H. ALBERT (Mannheim, BRD) – D. FØLLESDAL (Oslo, Norway) – P. GOCHET (Liège, Belgique) – J. HINTIKKA (Helsinki, Finland) – J. KIM (Ann Arbor, Mich., USA) – F. v. KUTSCHERA (Regensburg, BRD) – K. LAMBERT (Irvine, Cal., USA) – K. LEHRER (Tucson, Arizona, USA) – J. L. MACKIE (Oxford, England) – G. PATZIG (Göttingen, BRD) – M. PRZELECKI (Warsaw, Poland) – W. STEGMÜLLER (Munich, BRD) – P. WEINGARTNER (Salzburg, Austria)

Review Editor and Editorial Assistant: Wolfgang L. GOMBOCZ, Reinhard FABIAN

A major objective of "GRAZER PHILOSOPHISCHE STUDIEN" is to stimulate and to strengthen communications between American and European analytical philosophy. "GRAZER PHILOSOPHISCHE STUDIEN" are published by RODOPI of Amsterdam. One or two volumes (from time to time on special topics) of at least 200 pages will appear yearly. Correspondence concerning editorial matters, manuscripts for publication (in duplicate) and books submitted for review should be sent to Prof. Dr. Rudolf HALLER, Philosophisches Institut, Universität Graz, A-8010 Graz, Austria. Business correspondence and subscriptions please address to EDITIONS RODOPI N.V., Keizersgracht 302-304, Amsterdam, Holland.

Out of the contents 1 (1975) and 2 (1976)

Franz BRENTANO: Was an Reid zu loben. Über die Philosophie von Thomas Reid (Aus dem Nachlaß)

Roderick M. CHISHOLM: Individuation. Some Thomistic Questions and Answers

Joseph MAROLIS: Knowledge and Belief; Facts and Propositions

Eike von SAVIGNY: Das normative Fundament der Sprache: Ja und Aber

Further contributions and review articles by John B. BACON, Wilhelm K. ESSLER, Reinhardt GROSSMAN, Göran HERMEREN, Reinhard KAMITZ, Stephan KÖRNER, Keith LEHRER, Hans LENK, Ota WEINBERGER, and others.

Collected Papers on Epistemology, Philosophy of Science and History of Philosophy

by **Wolfgang Stegmüller**

Synthese Library 91
Complete set of 2 volumes
1977, approx. 544 pp.
Cloth Dfl. 145,-/US \$56.00

In these 2 volumes English translations of the author's main contributions to philosophy in the period 1956-1975 are published.

Three articles contain rational reconstructions of historical themes. In the first one results of W. V. Quine and N. Goodman are used to give a modern account of the *History of Universals*. In the second treatise *Kant's Philosophy of Science* is interpreted as an unparalleled attempt to solve the 'riddle of experience'. In the third article the basic ideas of *Wittgenstein's Picture Theory* are rendered precise with the help of an intensional variant of Tarski's concept of a relational system.

Among the systematic papers one on *Phenomenalism* gives a coherent treatment of all the problems which build an unsurmountable wall to establish the phenomenalist program.

Two papers on *Induction* deal in a new way with modern attempts to answer Hume's challenge. It is argued that the so-called problem of induction is to be replaced by successor-problems. Only some of them belong to the area of theoretical reasoning while the others fall into the field of practical reason. It is shown, in particular, that Carnap's inductive logic must be reinterpreted as belonging to the second domain. Part of this claim is substantiated by an imaginary dialogue between D. Hume and R. Carnap. In view of this result the Carnap-Popper controversy becomes completely pointless.

The essay on *Logical Understanding of the Dynamics of Theories* starts with an intuitive account of J. D. Sneed's non-statement view of theories, followed by a rational reconstruction of T. S. Kuhn's conception of a normal science. A simple solution of the Kuhn-Popper controversy is offered. Since Sneed's concept of theory-reduction allows the comparison of theories with completely different conceptual superstructures it is argued that the only remaining 'rationality gap' in Kuhn's account of scientific revolutions can be closed with the help of this concept. Besides, the new framework allows a precise and simple explication of Lakatos' conception of a research program.

The remaining articles deal with one of the main problems in the philosophy of humanities, i.e. the *Circle of Understanding*, and with topical problems in the philosophy of logic and philosophy of science.



D. Reidel Publishing Company

P.O. Box 17, Dordrecht-Holland
Lincoln Building, 160 Old Derby Street, Hingham, Mass. 02043 - U.S.A.

PRINTED IN ENGLAND

by William Clowes & Sons, Limited, London, Beccles and Colchester

AMERICAN PHILOSOPHICAL QUARTERLY

PRIZE ESSAY FOR 1978

Topic:

THE ETHICS OF BELIEF:
WHY SHOULD I BE RATIONAL?

Deadline:

JANUARY 15, 1978

Prize:

\$100 and publication in the OCTOBER 1978 issue of the APQ.

Conditions:

Only current subscribers to the APQ are eligible to make submissions. These must be written in English. They should be presented in the standard APQ typescript format and must not exceed 7,000 words in length. Entries will be judged by a panel of readers appointed by the editor of the APQ. The right of declining to make any award is reserved if no submissions of suitable quality are received.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Annette Baier
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
William T. Blackstone
Myles Brand
Robert Butts
Roderick M. Chisholm

L. Jonathan Cohen
James Collins
Antonio S. Cua
Arthur C. Danto
Richard M. Gale
Peter Thomas Geach
Marjorie Grene
Adolf Grünbaum
Rom Harré
Hugues Leblanc

Keith Lehrer
Gerald J. Massey
George Mavrodes
Alvin Plantinga
Jerome Schneewind
George A. Schrader
Wilfrid Sellars
Ernest Sosa
Richard Taylor
John W. Yolton

VOLUME 14/NUMBER 4

OCTOBER 1977

CONTENTS

- | | | | |
|---|-----|---|-----|
| I. J. P. DAY: <i>Threats, Offers, Law, Opinion and Liberty</i> | 252 | V. CHARLES HARTSHORNE: <i>The Neglect of Relative Predicates in Modern Philosophy</i> | 309 |
| II. EDMOND WRIGHT: <i>Perception: A New Theory</i> | 273 | | |
| III. JAMES W. CORNMAN: <i>Foundational versus Nonfoundational Theories of Empirical Justification</i> | 287 | VI. LAWRENCE H. DAVIS: <i>Prisoners, Paradox, and Rationality</i> | 319 |
| IV. PETER VAN INWAGEN: <i>Creatures of Fiction</i> | 299 | VII. MYLES BRAND: <i>Identity Conditions for Events</i> | 329 |

© AMERICAN PHILOSOPHICAL QUARTERLY, 1977

PUBLISHED IN OXFORD BY BASIL BLACKWELL

ISSN 0003-0481

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes" (short or long). Moreover the editor will only consider papers whose authors certify that while under consideration with us, they will not be submitted elsewhere.

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 12,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$18.00, increasing by \$2.00 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$12.00, increasing by \$2.00 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4.00 per group of 50.

SUBSCRIPTIONS

The subscription for 1977 is \$28.00 for institutions, \$12.00 to individuals. Single issues will be \$8.00 to all purchasers. For Monograph prices, see below. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Journals Department, Basil Blackwell & Mott Ltd., 108 Cowley Road, Oxford OX4 1JF, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The complete series is listed on the back cover. In 1977 these will be available at \$9.00 to institutions and non-subscribers, \$4.50 to individual subscribers. *Please note that it is no longer possible to supply the monographs free of charge to individual subscribers.*



I. THREATS, OFFERS, LAW, OPINION AND LIBERTY

J. P. DAY

1. Introduction

DO threats and offers curtail liberty? The doctors disagree over this question, which is a sure sign of a philosophical problem in need of treatment. The following summary shows the nature and extent of the disagreement. (1) Both do. (Aristotle, *Nic. Ethics*, tr. Ross, 1110a; Bay [1958], p. 93; Cassinelli [1966]; Frankfurt [1973].) (2) Threats do, but offers do not. (Benn and Weinstein [1971], p. 201; Hart and Honoré [1959]; Nozick [1972].) (3) Neither do. (Parent [1974 (2)]; Steiner [1975].) (See also Pennock ([1972], pp. 5f.).) Unsurprisingly, no one maintains that (4) offers do, but threats do not.

Unsurprisingly, too, this crude categorization needs qualification and amplification. Thus, as to (1), it is usually maintained that both *sometimes* do, namely, when the threat is strong or the offer is large. As to (1) and (2), it is usually contended that, when the threat is strong, the threatenee is not *fully* free, or has no *real* choice. As to (3), the position of Parent, e.g., is that threats do not diminish *social freedom*, though they do sometimes diminish *freedom of choice*, which he takes to be a different concept from social freedom (Parent [1974 (2)], p. 156).

The doctors disagree not only in what they say, but also in why they say it. For they are predominantly either (i) arguers or (ii) reporters. Thus, (i) Frankfurt, Nozick and Steiner offer argument in support of their positions. But, (ii), Aristotle follows his usual procedure of considering what we ordinarily say (for a start, at any rate); and so do Benn and Weinstein, Hart and Honoré, and Parent. However, this crude categorization, too, needs qualification and amplification. For, firstly, in conceptual analysis, the distinction between reporting and arguing is in any case not clear-cut. Secondly, whereas Aristotle and Parent claim to report truly what is ordinarily (or correctly) said in common speech, Hart and Honoré and Benn and Weinstein claim to report not only this, but also what is ordinarily said in the semi-technical discourse of *politics and law*. E.g., on politics, Hart

writes "(coercion) into obedience by the threat of legal punishment . . . rather than physical restrictions is what is normally meant in the discussion of political arrangements by restrictions on liberty" (Hart [1961], p. 21). While on the law, he points out, e.g., the following fact. Suppose that a highwayman (H) points a pistol at a traveller (T) and says to him "Your money or your life"; and that T hands over the money which he is carrying, which, however, happens to be not his own but his employer's. The common man would say that T's action was "not wholly voluntary," whereas the Anglo-American lawyer would say that it was done "without *mens rea*"; but both would say that it was "excusable" (Hart [1962], pp. 163f.).

I consider position (2) (above) to be the correct one, and my primary objective is to provide argument to show why it is so. My argument is different from Nozick's, however, and so will be found, I hope, to complement his valuable discussion. Finally, it may well be thought that this question forms only a small part of the general philosophical question, or questions, about liberty. We shall find, however, that this is not the case, and that answering it involves providing an elucidation of the idea of liberty in general; which is, indeed, my secondary and main objective in this essay.

2. Desire and Liberty

Steiner argues as follows to prove that neither threats nor offers, which he calls collectively "interventions," diminish liberty (Steiner [1975], p. 43). The effect of H's threat to T (above). "Either you will not have your money, or you will not have your life," is normally to change T's desire. Before the threat, T wants to keep his money, but after the threat he wants not to do so. Likewise with offers. H's conditional offer to T, "If you will give me your money, then I will give you my life," normally implies (in the popular, not in any logician's sense of "implies") "Either you will not have your money, or you will not have my life." The effect of this, again, may be to change T's desire to keep

his money into a desire not to do so. However, whether T desires to have his money or not is irrelevant to whether T is unfree to have his money or not. So the argument goes thus. (1) The effect of any intervention is normally to change the desire of *A* (the intervenee) to *D* into a desire to *-D*. (2) "*A* desires to *D*" and "*A* desires to *-D*" are both irrelevant to "*A* is unfree to *D*." Therefore, (3) No intervention curtails liberty.

I shall argue now, not that (3) is false (though I shall do this later [Sec. 3]), but rather that Steiner's argument fails to prove (3) true because one of its premisses is false. The false premiss is not, however, (2). For it is now generally agreed that, although the truth or falsity of "*A* desires to *D*" is relevant to the truth or falsity of "*A* feels unfree to *D*," it is not relevant to the truth or falsity of "*A* is unfree to *D*." This is what is wrong with, e.g., Mill's definition of liberty, "liberty consists in doing what one desires" (*On Liberty*, v). (See Friedman [1966]; Berlin [1969], pp. xxxviii ff.; Cassinelli [1966], p. 22; Day [1970], pp. 190 ff.; Oppenheim [1961], pp. 143 ff.; Steiner [1975], pp. 33 ff.). It is premiss (1) which is false. For, logically, there are four possibilities. T can (i) both have his money and have his life; (ii) not have his money but have his life; (iii) have his money but not have his life; (iv) both not have his money and not have his life. Happily for T, however, (iv) can be eliminated. This is because H's threat, "If you will not give me your money, then I will take your life" normally implies (in the popular, not any logician's sense of "implies") "If you will give me your money, then I will not take your life." Now, in fact, H's threat does not affect T's desire at all, since what T would like to do, both after the threat as well as before it, is (i). The effect of H's threat is rather to bring T to realise that, after the "dead-line" has been passed, he will be unable to do as he wants to do, namely, (i); and so to reflect whether he is less averse to doing (ii) than to doing (iii), or conversely; and to choose accordingly. Normally, he will choose alternative (ii). However, because the first conjunct of (ii) is "not to have his money," it *looks as though*, after and because of the threat, T wants not to have his money. This is what makes Steiner's argument seem plausible.

A possible obstacle to the realization that desire is irrelevant to liberty is the existence of the locution "*A* is free to *D* if he wants to." As Benn and Weinstein point out, the force of this is ironic, as in "Of course you are perfectly free to cut off your ears, if that is what you really want to do" (Benn

and Weinstein [1971], p. 195). Similarly, Hare notes the irony in "If you want to break your springs, go on driving as you are doing" (Hare [1952], p. 36). This locution brings out an important point about the relation of liberty to desire, which is that we do not normally speak of *A* being (un)free to *D* unless to *D* is the sort of thing that the average, reasonable man might want to do. However, a literalist might be misled by the locution into believing that the truth of "*A* desires to *D*" is a sufficient condition of the truth of "*A* is free to *D*". In fact, of course, "You are free to cut off your ears if you want to" no more entails "If you are not free to cut off your ears, then you do not want to" than "There is beer if you want it" entails "If there is no beer, then you do not want any."

A more serious cause of the false belief that the concept of liberty is linked to that of desire is the confusion of "voluntarily (or freely)" with "willingly (or gladly)." Hobbes, e.g., writes as follows respecting a case discussed by Aristotle (*l.c.*), from whom he doubtless took it. "...when a man throweth his goods into the Sea for fear the ship should sink, he doth it nevertheless very *willingly* ... It is therefore the action, of one that was *free*" (*Leviathan*, II, xxi, first emphasis added). This sounds plausible; yet, on consideration, it is plainly false. For, in the circumstances which Hobbes describes, the average man does not jettison his goods willingly (or gladly). On the contrary, he jettisons them very unwillingly (or reluctantly). On the other hand, he does jettison them voluntarily (or freely), since no one compels him to do so. Locke makes the same mistake as Hobbes. He considers the case where *A* is locked in a room where there is *B*, whom *A* wants very much to see. Consequently, *A* remains in the room very willingly; from which Locke infers, wrongly, that he does so voluntarily (*Essay*, II, xxi, 10; cp. Melden [1961], p. 219; O'Connor [1971], p. 77). Hobbes' example shows how *A* can *D* voluntarily but unwillingly. For an illustration of how *A* can *D* willingly but not voluntarily, consider the good soldier who dies executing an order to hold his position at all costs. He does not act voluntarily, because he cannot leave his position, since if he tries to do so, his superior will shoot him. Yet he acts willingly, for he believes that his cause is just, and that the execution of the order is necessary for the victory of his cause. However, "*A* *Ds* willingly" entails "*A* wants to *D*." So he who confuses "willingly (or gladly)" with "voluntarily (or freely)"

will believe, falsely, that the truth of "*A* desires to *D*" is a necessary condition of the truth of "*A* is free to *D*."

The latest type of desire (or will) theories of liberty is based on the concept of a second-order desire. According to Plamenatz and Dworkin, *A* *Ds* unfreely if and only if he *Ds* from a motive from which he desires not to act (Dworkin [1970]; Plamenatz [1938], p. 122; *cp.* Locke [1975], pp. 104ff.). However, if *H* kills *T* because he desires to have *T*'s money, although he desires not to have that base desire; it does not follow that *H* acts unfreely in killing *T*. "I did not want to want to do it" is no more an excuse than is "I did not want to do it." This sort of theory too breaks down on the fact that desire (or will) is irrelevant to liberty. (The special case in which *A* *Ds* unfreely because *B* arouses in *A* an irresistible desire to *D*, will be considered later [Subsec. 5.2].)

3. Threats, Offers and Liberty

H's threat to *T*, "Either you will not have your money, or you will not have your life," curtails *T*'s liberty because it makes *T* (about to be) unable to do something which he can now do, namely, (i) both have his money and also have his life (Sec. 2). (It is assumed that the threat is a normal, i.e. genuine one; i.e., that *H* sincerely intends to execute his threat and is not bluffing; that he is able to execute his threat and has not forgotten to load his pistol, and so forth.) It is necessary to say that *H*'s threat makes *T* "about to be" unable to do (i), because he will not actually be unable to do it until the time-limit has expired. With a normal threat like this one, the threatener always allows an interval between making it and executing it, so that the threatenee can reflect before deciding between (ii) and (iii). Similarly, with normal offers, the offerer allows the offeree time to think it over before deciding whether to accept it or to reject it.

The failure to understand why and how threats curtail liberty proceeds from concentrating attention on *T*'s simple action, keeping his money, rather than on the complex (i.e., conjunctive) action (i). The difficulty is that it is rightly recognized, on the one hand, that *H*'s threat curtails *T*'s liberty *somehow*; and yet, on the other hand, that it does not make him unfree to keep his money. The first step towards understanding why and how threats curtail liberty is to recognise that a threat (unlike an offer [Subsec. 6.1]) must be complex, and is best expressed by an alternation; and that any alternation negates some conjunction. In the Calculus of

Propositions, the formula $(\neg P \wedge \neg Q) / (P \& Q)$ is a tautology. (*Cp.* Pollock [1973], p. 235.) I think that this point may have been overlooked because philosophers of action tend to neglect the important subject of complex actions.

In the light of these considerations, it is possible to reply to an objection by Cassinelli to the view that threats diminish freedom (Cassinelli [1966], pp. 33f.). He considers the case where *A* murders *B* in a state in which there is a law against murder which is backed by a threat of punishment by death. The fact that *A* nevertheless did murder *B*, says Cassinelli, shows that the threat did not make him unfree to do so; for it is illogical to say that *A* was unfree to do something which he in fact did. The reply to this objection is that it is quite true, but that it commits the fallacy of *ignoratio elenchi*. For it does not show that *A* was not unable both to take *B*'s life and also to keep his own life which is how this threat in fact curtails *A*'s liberty (Subsec. 6.2).

But does not *H*'s complex (i.e., conditional or alternative) offer to *T*, "Either you will not have your money, or you will not have my life," similarly diminish *T*'s liberty by making him (about to be) unable to do something which he can now do, namely, both have his money and also have *H*'s life? It does not do so. For the crucial distinction between threats and offers is as follows. Before *H*'s threat, *T* does and therefore *can unconditionally* both have his money and also have his life; whereas after *H*'s threat *T* cannot do both these things. But before *H*'s offer, *T* *cannot unconditionally* both have his money and also have *H*'s life. For, in this context, "have *H*'s life" means "receive *H*'s life from *H*." And *T* cannot do this *unless and until H gives his life to T*. Consequently, *H*'s offer does not make *T* (about to be) unable to do something which he can unconditionally now do. *H*'s threat *deprives* *T* by making him (about to be) unable to do something which he can unconditionally now do; whereas *H*'s offer does not do this. This is the vital difference between threats and conditional offers, on which hinges the fact that threats curtail liberty whereas conditional offers do not, and to recognize it is to take the second step to understanding the relation of threats and offers to freedom. Furthermore, we shall see shortly that offers not only never decrease the liberty of the offeree, but also sometimes (though not always) increase it (Sec. 4). Threats, on the other hand, never increase the liberty of the threatenee. What has been said is true of any threat, regardless of its strength. *H*'s very weak threat, "Either you will give me your

money, or I will give you the rough edge of my tongue," also diminishes T's freedom, because it makes him (about to be) unable to do something which he is now doing and therefore can unconditionally now do, namely, both have his money and also have his ears secure from verbal assault. (Cp. Austin [1961]; Locke [1975]; Moore [1912]; Nowell-Smith [1954]; O'Connor [1971], pp. 26ff.)

4. *Ability and Liberty*

B decreases A's liberty by decreasing the number of alternative actions which A can perform. The effect of H's threat to T, "Your money or your life," is to eliminate (in the near future) the alternative of T's both keeping his money and also keeping his life. Similarly, T now has the alternatives of either going from London (L) to York (Y), or staying in L. If H imprisons T, he eliminates the former alternative, and thereby decreases T's freedom. So may not B increase A's freedom by increasing the number of alternative actions which A can perform? Some certainly think so. MacMurray, e.g., maintains this position as part of the following general thesis on liberty. Freedom is the *summum bonum*, and there are two main roads to it, the philosophical-religious high road and the scientific-technological low road. The former, Platonic-Stoic-Christian way is the way of diminishing our desires, and the latter, modern Western way is the way of increasing our powers. MacMurray thinks that the low road has turned out to be no road, and that we should do better to return to the high road (MacMurray [1949], I).

We have in effect already rejected MacMurray's thesis about the high road as resting on a misconception of the relation of liberty to desire (Sec. 2). If H makes T unable to go to Y by imprisoning him, then T will not cease to *be* unfree to go if he suppresses his desire to go; though he will admittedly cease to *feel* unfree to go if he does this. So T will not increase his freedom by decreasing his desires. But we must also reject MacMurray's thesis respecting the low road.

Two centuries ago, when T wanted to go from L to Y, he had no alternative but to go by road. One century ago, he had the alternatives of going either by road or not by road, but the latter alternative took the single form of going by rail. Today, however, the alternative of going not by road includes the two sub-alternatives of going by rail or going by air. The question is whether the inventors of railways and aircraft have made T

more free. It may be argued that they have, as follows.

When H releases T from prison, he adds to T's alternative of staying in L the alternative of going to Y, and thereby augments T's freedom. Similarly, when the inventor of aircraft adds to T's options of going to Y by road or by rail the further option of going by air, he too increases T's liberty. But the objection is that there is a crucial difference between the two cases. Before H released T from prison, T was *unfree* to go to Y; and this is why H's giving T the additional alternative of going to Y increases T's *freedom*. But before their inventor invented aircraft, T was not unfree, but *unable* to fly to Y; and this is why their inventor's giving T the additional alternative of going to Y by air increases T's *power*. Now liberty, so far from being identical with ability (or power), presupposes it. I.e., the truth of "A is able to D" is a necessary condition both of the truth and also of the falsity of "A is free to D" (Day [1970], p. 180). Consequently, it is neither true nor false that, one century ago, T was unfree to fly from L to Y. Rather, the question, whether he was unfree to do so or not, does not arise.

The case is similar with offers; for these, like inventions, often increase the number of alternative things that A can do (Benn and Weinstein [1971], p. 201). So do such offers augment liberty? The answer is implicit in what has just been said. If H first imprisons T in L, and then says to him "If you will give me your money, then you may go to Y," he increases T's liberty by giving him the alternative of going to Y; because before the offer T was *unfree* to go to Y. But if H says to T, as T is setting out for Y on foot, "If you will give me your money, then you may have my horse," then H does not increase T's liberty by giving him the extra option of going to Y on horseback; because before the offer T was not unfree, but unable to go to Y on horseback. In this case, the effect of H's offer is to increase T's power, not his liberty.

So the answer to the question posed at the beginning of this Section is that only sometimes does B increase A's freedom by increasing the number of alternative actions which A can perform. At the other times B thereby increases, not A's freedom, but his power.

One likely cause of the confusion of liberty with ability is the fact that the same word, "can," is used for both concepts. The point of the statement, "Today, T can fly from L to Y," may be either (1) that today we have aircraft, whereas one century

ago we had not (ability); or (2) that today there is no governmental prohibition on civil avation, whereas yesterday there was (liberty).

Moore's theory fails through not distinguishing unfreedom from inability. According to Moore, *A* is unfree when and only when he cannot -*D* if he chooses (Moore [1912]; Austin [1961]). Yet when *A* is so sick that he cannot leave his bed if he chooses, he does not stay in his bed unfreely. His illness diminishes his power, not his liberty. (Cp. Hobbes, *l.c.*).

5. Conditions of Unfreedom

Generally speaking, it is not disputed that the truth of "*B* makes *A* unable to *D*" is a sufficient condition of the truth of "*A* is unfree to *D*." (However, we shall have to notice an important qualification of this statement shortly [Subsec. 5.5.].) Consequently, in saying that *H*'s threat to *T*, "Your money or your life," curtails *T*'s liberty because it makes *T* (about to be) unable both to have his money and also to have his life, I am saying nothing controversial. I am saying that the way in which this threat curtails *T*'s liberty is not essentially different from the way in which *H* curtails *T*'s liberty to go from *L* to *Y* when he imprisons him. However, some maintain that threats and offers diminish freedom in ways that are different from this way.

5.1. Ineligible Alternatives

We have seen that Benn and Weinstein hold that threats curtail liberty but that offers do not (Sec. 1.). They contend that threats do this by making some alternative action ineligible. "Ineligible" is an evaluative term, and the concept of an ineligible alternative is relative to the average man, so that "an ineligible alternative" means "an alternative which it is unreasonable for the average man to choose." Only a strong threat, such as "Your money or your life," confronts *T* with an alternative that it is unreasonable for the average man to choose, namely, to keep his money but to lose his life. A weak threat, such as "Either you will give me your money, or I will give you the rough edge of my tongue," does not diminish freedom, because it would not be unreasonable for the average man to choose to keep his money but to forfeit the security of his ears from verbal assault. On this account, therefore, *H*'s strong threat curtails *T*'s liberty not (or at any rate, not only) by making it (about to be) impossible for him both to keep his money and also to keep his life, but (or

at any rate, but also) by making it unreasonable for him to keep his money. *H* does this by attaching to *T*'s keeping his money the very high price of losing his life; for in this situation, to choose to keep his money is in effect to choose both to keep his money and also to lose his life. On this view, therefore, there are two ways in which *B* can make *A* unfree, namely, (1) by making some alternative impossible for *A*, and (2) by making some alternative ineligible for *A*.

This theory seems to me to be open to the following objections. Firstly, it involves Cassinelli's paradox (Sec. 3). For suppose that *T* is unreasonable; i.e., that he keeps his money and loses his life. Then, on this theory, *T* does a thing which he is unfree to do. But this is illogical.

Secondly, suppose that *H* holds out to a female *T* a large sack full of money and says to her "You may have all this money." Then *H* presents *T* with two alternatives, viz., either to accept his offer or to reject it, the latter of which is very ineligible. On this theory, therefore, *T* is unfree to decline the money. But this contradicts its authors' thesis that offers do not diminish freedom.

Thirdly, suppose that the sum of money which *T* is carrying is very large and furthermore all that he possesses; and that *H*, without showing a gun, says to him imploringly "Please give me your money." Then *H* presents *T* with two alternatives, viz., either to grant *H*'s request or to refuse it, the former of which is very ineligible. On this theory, therefore, *T* is unfree to hand over his money. But this conflicts with the received view that requests do not diminish freedom.

By attaching to *T*'s keeping his money the price of losing his life, *H* deters (or attempts to deter) *T* from keeping his money. But *B* does not make *A* unfree to *D* by deterring him from *D*ing. To do this, *B* must prevent *A* from *D*ing. The ineligible alternatives theory is so plausible precisely because men tend not to distinguish between deterrence and prevention. However, we shall see later that this distinction is crucial for the analysis of the concept of liberty (Subsecs. 6.2., 6.3).

5.2. Irresistible Motives

We have also seen that Frankfurt maintains that both threats and offers diminish freedom (Sec. 1.). He argues that, in the case of strong threats, *B* makes *A* unfree to *D* by arousing in him a motive "so powerful that he cannot prevent it from leading him to submit to the threat" (Frankfurt [1973], p. 77). Similarly, in the case of attractive offers, *B*

makes *A* unfree to *D* by arousing in him an irresistible desire to gain the offered benefit.

This theory seems to me to be open to the following objection. Suppose that, when *H* says to *T* "Your money or your life," *H*'s mien is so terrifying and *T*'s disposition is so timid that *H*'s action arouses in *T* an intense fear amounting to panic, in consequence of which *T* hands over his money "automatically." This is perfectly possible; but it is not a case of pure threatening. For, although *H* *menaces* *T*, he also *intimidates* him; and *T* does not *comply with H's threat* at all, but is cowed into doing as he does. Although threatening and intimidating are usually not distinguished, they are, I think, quite different concepts. In the field of labor-relations, e.g., it is surely no less important (and difficult) to distinguish threatening from intimidating than it is to distinguish threatening from warning (see Nozick [1972], pp. 120ff.). Paradoxical as it may seem, the paradigm threat situation is one from which emotion is wholly absent; i.e., one in which *H* is quite unalarming and/or *T* is quite fearless, and in which *T* chooses to keep his life and lose his money, rather than conversely, from as dispassionate a consideration of his *interest* as when he decides to spend a legacy on a new car rather than on new furniture. The failure to distinguish between threatening and intimidating is explained by an obvious overlap between the concepts. Very often, when *B* threatens *A*, he also (not necessarily intentionally) intimidates him; and when *A* complies with *B*'s threat, he is very often moved by fear as well as by interest.

Similarly with offers. Suppose that *H* holds out to a female *T* a large sack full of money and says to her "You may have all this money if you will give me your hand in marriage," and that *T* is both avaricious and destitute. Then *H*'s action may arouse in *T* a desire so intense that she "instinctively" snatches the sack with one hand and gives her other hand to *H*. However, this is not a case of pure offering. For, although *H* *makes T an offer*, he also *tempts* her; and *T* does not *accept H's offer* at all, but is lured into doing as she does. Here again, the paradigm offer situation is one from which emotion is wholly absent, and in which *A* decides to accept or to decline *B*'s offer by a purely rational calculation of his interests. But here, too, there is an obvious overlap between offering and tempting. Very often, when *B* makes *A* an offer, he also (not necessarily intentionally) tempts him; and when *A* accepts *B*'s offer, he is very often moved by desire as well as by interest. Moreover, although not all

offering involves tempting, all tempting does involve offering.

I agree with Frankfurt, then, that *B* can make *A* unfree to *D* by making him unable to *D* through arousing in him either an irresistible fear of *Ding*, or an irresistible desire of *-Ding*. But I do not agree that this shows that and how threats and offers curtail liberty, because what have just been described are extreme intimidation and extreme temptation; and intimidating is not identical with threatening, nor is tempting identical with offering. However, the fact that *B* can simultaneously both threaten *A* and intimidate him, or both make *A* an offer and tempt him, brings out again the importance of complex (conjunctive) actions (Secs. 3.6). We describe these complex actions by such locutions as "*B* made *A* a very tempting offer."

5.3. Psychological Unfreedom

Some say that, in the cases of extreme intimidation and extreme temptation just discussed (Subsec. 5.2), *H* coerces *T* psychologically. But why "psychologically"? If *T*, moved by extreme fear, hands over his money "automatically", or, moved by extreme desire, snatches *H*'s money "instinctively", then we describe what *T* does as "purely physical reactions." Moreover, it is precisely and only in these extreme cases where there are purely physical reactions that *T* is coerced. When *T* is only moderately frightened or tempted, he is not unfree. In any case, the "physically/psychologically" distinction is not clear-cut. If *H* uses post-hypnotic suggestion to make *T* unable to walk from *L* to *Y*, does *H* coerce *T* psychologically or physically? The important point is that, for the purposes of this inquiry, it does not matter. All that matters is that *H* *makes T unable* to walk from *L* to *Y*; whether "physically" or "psychologically" is immaterial.

Cassinelli says that the kleptomaniac (*K*) is psychologically unfree to refrain from stealing because his irresistible desire to steal makes him incapable of not stealing (Cassinelli [1966], pp. 39f.). But this is incorrect, for an important reason. Unfreedom is essentially an *interpersonal* affair, in which *B* *makes A unable* to *D*. But *K* is not *made* unable to refrain from stealing by another person; he just *is* unable to do so. Or, if he is made unable to do so, it is not by another *person*, but by his own irresistible *desire*. This is the conception of freedom as an *intrapersonal* relation between one part of one man's mind and another part of it. This vital distinction is clearly revealed in the language of the psychologists, who call *K*'s thefts not *com-*

pulsory, but *compulsive*. "Compulsory," or "compelled," are ellipses of "compelled by *B*"; whereas "compulsive" does not thus implicitly connote the existence of a compeller. This, again, is why *K*'s case is quite different from that discussed above (Subsec. 5.2) in which *T* is tempted to take *H*'s money; for in that case *T* is tempted *by H*.

Bay makes extensive use of the concept of psychological unfreedom. He derives it from Green's and Bosanquet's notion of "positive" freedom as self-realization, and brings it up to date with the findings of modern psychologists, especially those of Fromm. "Positive freedom . . . is . . . the full realization of the individual's potentialities. . . ." For Bay, the unfree man is the psychotic or severe neurotic, and Bay's model of him is the authoritarian personality (*A*), whose defensiveness is caused by fear.

But the case of *A* is on all fours with that of *K*, just discussed. *A* is not unfree, because he is not made unable to realize his potentialities by *B*; he simply is unable to do so. (Cp. Pennock [1972], p. 6.) Or, if he is made unable to do so, it is not by another person, but by his own fears.

Bay recognizes the intrapersonal character of his "psychological unfreedom," and contrasts it with two interpersonal concepts, "social unfreedom" and "potential unfreedom," the latter of which I shall consider shortly (Subsec. 5.4). He also recognizes that the Idealists' intrapersonal conception of liberty originates with Plato, in whose terms *A* is "a slave to his fears" because "the better part (of his soul) is overwhelmed by the worse" (*Republic*, 430-31, tr. Cornford). But then, as Green and Bosanquet stress, this notion of Plato's rests on nothing more substantial than the most famous of all his political *analogies*, viz., that which he draws between the good man and the good society—"a man is just in the same way that a state (is) just" (*l.c.*, 441; Green [1937], pp. 3f., 9f.; Bosanquet [1923], pp. 128ff.). We must resist the lure of Plato's master/slave trope, and insist that liberty is necessarily interpersonal. Speaking strictly and correctly, "*A* is unfree to *D*" entails "*B* makes *A* unable to *D*," where the variables *A* and *B* range over agents and the variable *D* ranges over simple or complex actions.

Plato's notion of intrapersonal freedom is the main source of the thesis that liberty is virtue. For in his view the unfree man, who is a slave to his passions, is also the bad (or "unjust") man; whereas the free man, who is master of himself (i.e., of his passions), is the good (or "just") man.

5.4. Potential Unfreedom

Bay's potentially unfree man is the manipulated man (*M*), who is so called because the open or hidden persuaders who manipulate him "hamper (his) potential behaviour, not his actual behaviour." Bay finds the source of this notion of unfreedom in Rousseau:—"If it is good to know how to deal with men as they are, it is much better to make them what there is need that they should be. The most absolute authority is that which penetrates into a man's inmost being, and concerns itself no less with his *will* than with his *actions*" (*A Discourse on Political Economy*, tr. Cole, emphases added). Skinner gives a contemporary expression to the same thought:—" . . . a system of slavery so well designed that it does not breed revolt is the real threat (to freedom)" (Skinner [1971], p. 40).

However, *M*'s plight consists not so much in the fact that he is unfree as in the fact that when he is unfree he fails to feel unfree. A dictator might make all the citizens unable to travel abroad by preventing them from embarking in ships or aircraft, and simultaneously eliminate their desire to do so by introducing subliminal propaganda into TV programs to the effect that foreign travel is unpatriotic and narrows the mind (Benn [1967]; Day [1960], p. 62). Then the citizens would *be* unfree to go abroad, but would not *feel* unfree to go. The prevention (or arousal) of *M*'s desires, of which Rousseau speaks, by means of rhetoric and persuasion, advertising and propaganda, etc., has everything to do with making *M* not feel unfree, but nothing to do with making him be unfree. That is a matter of the prevention of his actions.

Most misunderstandings about liberty proceed from a misconception of the relations of liberty to desire (or will) (Sec. 2) and/or to ability (or power) (Sec. 4). Bay's concepts "potential freedom" and "psychological freedom" illustrate this truth. What he calls "social freedom" is in fact freedom *sans phrase*, and we are now ready to elucidate this potent idea.

5.5. Final Analysis of Unfreedom

Our inquiry into the conditions of unfreedom has reached the following position. Critically, we have rejected the theses that *A* is unfree to *D* when he is incapable of *D*ing (Sec. 4); or when he acts unreasonably in *D*ing (Subsec. 5.1), or when he acts immorally in *D*ing (Subsec. 5.3); or when he does not feel unfree in *D*ing (Subsec. 5.4).

There is one further thesis to be dismissed. This is the position that *A* is unfree to *D* if he acts pre-

dictably in *Ding*. For I accept what O'Connor [1971] calls "the consistency hypothesis", i.e., Hobbes' contention that "*Liberty and Necessity are consistent*" (l.c.: cp. Hume, *An Enquiry concerning Human Understanding*, VIII; Mill, *A System of Logic*, VI, ii). When H points his pistol at T and says "Your money or your life," we can predict that T will almost certainly hand over his money, because we know that this is what almost everybody does in such a case. But it neither follows nor is true that T is unfree to keep his money in this situation; as he would be if H forced him to hand it over, or cowed him into handing it over. In the threat situation, T's action in handing over his money is both predictable and also free. (Cp. Ayer [1959]; Fain [1958].)

Constructively, we have found that *A* is unfree to *D* when *B* makes it impossible for him to *D*. E.g., T is unfree to keep his money (1) when H forces him to hand it over, or (2) when H cows T into handing it over. Again, (3) T is (about to be) unfree both to keep his money and also to keep his life when H makes T (about to be) unable to do these two things (the threat situation). Consequently, it looks very much as if *B*'s making *A* unable to *D* is the unique sufficient condition of *A*'s being unfree to *D*. I shall claim accordingly that, subject to a qualification to be made shortly, the truth of "*B* makes *A* unable to *D*" is the sufficient and necessary condition of the truth of "*A* is unfree to *D*." So we arrive, in the end, at the formulation of Bentham:— "... coercion, which (liberty) is the absence of . . ." (*Of Laws in General*; cp. Friedman [1966]). Or, in a contemporary version, "An individual is unfree if, and only if, his doing of any action is rendered impossible by the action of another individual" (Steiner [1975], p. 33).

However, two further clarifications are called for, both of which concern the relations of liberty to ability. Firstly, I pointed out earlier that liberty presupposes ability; i.e., that the truth of "*A* is able to *D*" is a necessary condition both of the truth and also of the falsity of "*A* is unfree to *D*" (Sec. 4). But since "*A* is unfree to *D*" is materially equivalent to "*B* makes *A* unable to *D*," it follows that the last formula also presupposes "*A* is able to *D*." And this is indeed so. H can make T unable to walk from L to Y, because that is a thing which T is able to do. But H cannot make T unable to walk from L to New York, because that is a thing which T is unable to do anyway. If *A* is unable to *D*, the question of his being made unable to *D* cannot arise.

Secondly, although the truth of "*B* makes *A*

unable to *D*" is a necessary condition of the truth of "*A* is unfree to *D*," the truth of the former is not strictly a sufficient condition of the truth of the latter, for an interesting reason. Consider the following three ways in which H can make T unable to walk from L to Y: (1) by imprisoning him; (2) by cutting off his legs; (3) by breaking his left ankle. In case (1), we have no hesitation in saying that T is unfree to walk from L to Y; it is a paradigm case of unfreedom. In case (2), which is also a paradigm, we have no hesitation in saying that T is unable, not unfree, to walk from L to Y. In case (3), we perhaps hesitate, but are prepared to allow that T is unfree, not merely unable, to walk from L to Y. Plainly, the difference resides in whether the inability is irretrievable or not. We do not talk of *A* being unfree to *D* unless *B* has made *A* unable to *D* in such a way that it is possible for *A* to be again able to *D*. T will again be able to walk to Y if he is released from prison, or when his ankle mends; but he will never be able to do so if his legs have been cut off. This is why, even when H imprisons T for the term of his natural life, we still call T "unfree." For we think that there is always a chance that he will escape or be rescued, or that H will relent. Let us therefore speak of *B* making *A* "(ir)retrievably" unable to *D*.

We have seen that not all increases in the number of alternative actions open to *A* augment his liberty, since some such increases augment, not *A*'s liberty, but his power (Sec. 4). Similarly, we now see that not all decreases in the number of alternative actions open to *A* curtail his liberty. For when *B* inflicts an irretrievable disability on *A*, he curtails, not *A*'s liberty, but his power.

The extreme irretrievable disability that *B* can inflict on *A* is death. Consequently, if H kills T, he makes him not unfree, but unable to walk from L to Y. However, *B*'s actually killing *A* must be distinguished from *B*'s being disposed to kill *A* under certain conditions, one (but not the only) case of which is *B*'s threatening to kill *A*. The good soldier is unfree to quit his post, not because his superior has shot him, but because his superior will shoot him if he tries to do so, so that he is unable to do so (Sec. 2). And *A*'s liberty is curtailed by a law against murder which is backed by a threat of punishment by death, not because his sovereign has killed him, but because his sovereign's threat makes him unable to both take *B*'s life and also to keep his own life (Sec. 3).

Finally, then, the sufficient and necessary condition of the truth of "*A* is unfree to *D*" is the truth

of "*B* makes *A* retrievably unable to *D* by *Eing A*." E.g., "H makes T retrievably unable to walk from L to Y by imprisoning T." The presence of the first agent-variable, *B*, brings out the essentially interpersonal character of freedom. The presence of the second action-variable, *E*, stresses the important truth pointed out by Bay, that "it is the *means* and not the *ends* of power exercise that determine the extent to which . . . freedom is interfered with" (Bay [1958], p. 91, emphases added). The presence of the modifier "retrievably" meets the point made in the preceding paragraph, and illustrates again the truth of Bay's contention that it makes a difference *how B* renders *A* unable to *D*. So, whenever any assertion is made about liberty, it is always possible to discover exactly what (if anything) is being asserted by seeking values for the four variables in the above formula, in order to obtain a clear statement which is either true or false. (Cp. MacCallum [1972].).

6. *The Logic of the Will and the Philosophy of Law*

The position which I have maintained thus far is essentially Hobbist. It will be fruitful to introduce now some leading ideas of Bentham's. Bentham is the creator of the imperative theory of law, according to which a law is a command to a subject that is declarative of his sovereign's will (*Of Laws in General*, i, 1). But more, he is the creator of the philosophy of law itself, since, as Mill says, "he found the philosophy of law a chaos, he left it a science" ("Bentham," *Dissertations and Discussions*, I). More yet, he is the creator of what he calls "the logic of the will," which is called today "deontic logic." (See Hart [1971], pp. 60ff.; McLaughlin [1973], pp. 215f.). So far, I have analysed the concept of liberty (or freedom) by exhibiting its relations to the four concepts of ability (or power), desire (or will), threat and offer. I shall now complete my analysis of it by showing its connexions with Bentham's three cardinal concepts, order (or command), permission, and law (or mandate). (Sec. 1, end.)

6.1. *Influence and Liberty*

I use "to influence" in a wide sense, such that "*B* influences *A* to *D*" is equivalent to "*B* either induces, or tries to induce *A* to *D*." For the purposes of this inquiry, the most important distinction between the different ways in which *B* can impose his will on *A* is that between coercive influence and non-coercive influence. *B* compels (or coerces) *A* to *D* if and only if *B* makes *A* retrievably unable to

-*D* by *Eing A*. E.g., H compels T to remain in L by making him retrievably unable to leave L through imprisoning him. Hence, *A* is unfree to *D* if and only if *B* compels *A* to -*D*. In other words, as Bentham says, freedom is the absence of compulsion (Subsec. 5.5).

(A) Coercive modes of influence are: (1) forcing, (2) threatening, (3) extreme intimidation, (4) extreme temptation, (5) extreme domination, and (6) extreme provocation. (B) Non-coercive modes of influence are: (7) ordering, (8) requesting, (9) offering, (10) persuading, (11) advising, (12) warning, (13) praising, (14) blaming, (15) rewarding, (16) punishing, (17) deterring. In (A), but not in (B), *B* is said in the language of the law to "cause" (or "make") *A* to *D* (Hart and Honore [1959], pp. 71, 173).

As to (2), the main difference between threatening and the other five coercive modes of influence is that, whereas in the latter *B* makes *A* unable to *D*, in the former *B* makes *A* unable both to *D* and also to *E*.

(5) and (6) resemble (3) and (4). In (5), *B* bullies *A* into *Ding*, since *B*'s ascendancy over *A* is so great that *A* is unable to -*D*. (6) includes extreme incitement, instigation and exhortation. E.g., *B* taunts *A* so grossly that *A* strikes *B* in purely reflex retaliation.

In the law, (4) and (6) are treated as calling for mitigation of penalty, but not for excuse from penalty (Hart [1962], pp. 168ff.). Yet it is hard to see any justification for this. It seems to me that (3), (4), (5) and (6) render *A* unfree, so that he ought to be excused punishment. Naturally, this is not to deny the difficulty of establishing in any particular case whether *B*'s temptation, or provocation, or etc. of *A* was in truth so strong as to make *A* unable to -*D*. Since we usually lack the relevant information about the individual *A*, we fall back on the concept of the average man. Aristotle, e.g., speaks of pardoning (i.e., excusing) *A* "when (*A*) does what he ought not under pressure which overstrains human nature" (*l.c.*). Aristotle's "human nature" is plainly the statistical concept, "the average man." But this difficulty arises whenever *B* is alleged to make *A* unable to *D*. Thus, does the locked door make the imprisoned T unable to go to Y? If T is a man of average strength, yes; but if he is exceptionally strong, no. So, unless we know that T is exceptionally strong, we deem him unable to go to Y (cp. Benn [1967], pp. 267f.).

As to (9), all offers (or invitations) are permissions, but not conversely. E.g. respectively (i) H

says to T "You may have all this money if you will give me your hand in marriage," and (ii) an employer says to his secretary "You may go if you have finished your work." Permissions of class (i) are influences, but permissions of class (ii) are not influences. We have seen that the two main differences between (9) and (2) are first, that threats coerce whereas offers do not coerce; and secondly, that offers can be either simple or complex, but threats can only be complex (Sec. 3). Examples of simple and complex offers are respectively "You may have all this money" ("no strings attached"), and (i), above.

(1), (10) and (17) differ from the other fourteen verbs under consideration in being, like "induce," what Ryle calls "achievement verbs" (Ryle [1949], pp. 149ff.). I.e., it is illogical to say "*B* forced *A* to *D*, but *A* did not *D*." The fact that, if *B* persuades *A* to *D*, then *A* cannot -*D*, tends to create the illusion that persuasion is a compulsive type of influence; and similarly with deterrence. Again, most of these verbs fall naturally into groups. Thus, dominating, intimidating, provoking and tempting are members of the same family. And requesting goes naturally with ordering, as does offering with persuading, and warning with advising. Also praising goes with rewarding, as blaming does with punishing.

Just as there are complex (conjunctive) actions which involve both threatening and intimidating, and both offering and tempting; so there are other sorts of complex actions which involve others of the seventeen sorts of actions that we are talking about (Subsec. 5.2). E.g., *B* can both order *A* to *D* and bully *A* into *Ding*; and *B* can both persuade *A* to *D* and provoke *A* into *Ding*. We describe these types of complex actions by such locutions as "*B* roared at *A* an order to *D*," and "*B* needled *A* into *Ding*." Consequently, just as it may be hard to distinguish offering from tempting, threatening from intimidating, and threatening from warning; so it may also be difficult to discriminate between ordering and bullying, persuading and provoking.

Again, one can combine an offer with a threat, as in "Either you will take my money, or I shall take your life." For these combinations, I shall borrow from Steiner the name "throffer." (Steiner himself, however, gives this name to more complex combinations, typified by "Either you will take *A*'s life and take my money, or I will take your life.") Throffers coerce, not *qua* offers (which are non-coercive), but *qua* threats (which are coercive). An important type of hybrid is exemplified by "Give

me your money! or I shall take your life," the obvious name for which is "thrororder." Steiner, indeed, gives this as an example of a threat; but this is not so, since the corresponding pure threat is, as we have seen, "Either you will give me your money, or I shall take your life" (Sec. 3). Thrororders too coerce *qua* threats, not *qua* orders. For *H*'s pure order to *T*, "Give me your money!," does not make *T* unable to keep his money, or indeed unable to do anything else. The request, "Please give me your money," is similarly noncoercive (Subsec. 5.1). But compare with this pure request the coercive and ironic threquest (for the use of gentlemanly *H*s), "Pray give me your money, or I shall take your life."

The fact that pure orders do not compel raises a doubt about whether there are such things, since they seem to be pointless. So may not what seem to be pure orders always be really either orders *implicitly* backed by threats, or *implicitly* thrororders? The answer to this question is, No. Speaking strictly, Bradley is right in saying that "threat is not of the essence of command; command need not imply the holding forth . . . of consequences" (Bradley [1927], p. 207). But Bradley's thesis is contradicted by the antithesis that all imperatives are "hypothetical" (in Kant's sense) and reducible to indicatives. According to this antithesis, "Let *A* *D*!" is always an ellipsis of "If *A* will not *D*, then it will be the worse for *A*" (or of "Either *A* will *D*, or it will be the worse for *A*") (Hare [1952], pp. 7ff., 33ff.). However, this antithesis differs importantly from the view that all orders are either (a) orders explicitly or implicitly backed by threats, or (b) explicit or implicit thrororders. (a) is illustrated by "Give me your money! For either you will give me your money, or I will take your life." For, on this last view, orders are certainly not reducible to *indicatives*. This is because the second alternant in the alternation (α) "Give me your money! or I shall take your life," and the consequent in the conditional (β) "If you will not give me your money, then I shall take your life" is not *descriptive*. It is rather a declaration of intention, and so what J. L. Austin calls *commissive*, because it commits him who says it to a certain course of action (Austin [1962], pp. 40ff., 156ff.). (β) is not a conditional prediction like (γ) "If you will not give me your money, then I shall be less rich than my brother." Hence the impossibility of a truth-functional treatment of threats, since commissives are neither true nor false. I mentioned earlier the difficulty and importance of distinguishing warnings from threats (Subsec.

5.2). Thus, if H says to T (δ) "If you will not give me your money, then I shall take my life," is that a threat or a warning? Without further information, there is no telling. E.g., if H knows that T desires H to take H's life, then (δ) is neither a threat nor a warning.

Economists sometimes distinguish "demand" from "effective demand," and a parallel distinction between "command" and "effective command" is helpful here. Nevertheless, it is false that all effective commands are either (a) orders explicitly or implicitly backed by threats, or (b) explicit or implicit orders. For a religious leader might order his disciples to give all their possessions to the poor, and the order might prove effective, neither because of any threat made or implied by the leader, nor because of his power, but because of his authority over his disciples. (Cf. Broadie [1972], p. 190.) But such cases are infrequent, and we must conclude, firstly, that effective orders are generally either of type (a) or of type (b); and secondly, that effective orders are coercive, not because they are orders, but because they involve threats. Finally, orders, like offers, can be either simple or complex, and complex in either the conjunctive or alternative modes. Examples of the two types of complex orders are respectively "Give me your money and give me your life!" and "Give me your money or give me your life!" An example of a conjunctive offer is "You may have my money and you may have my life."

As for (13)–(16), praise and blame, reward and punishment are of course the modes of influence to which moral philosophers have given most attention. I shall consider shortly the relations of the last two to threats, offers, law, opinion and liberty (Subsecs. 6.2, 6.3).

The seventeen modes of influence which I have discussed are the common ones. But there are others. One unusual one is post-hypnotic suggestion, which, as we have seen, is coercive (Subsec. 5.3). Ordinary suggestion, on the other hand, is often not a mode of influence at all; but when it is, it is non-coercive, and merges in persuasion. There are other such mergers. Thus, I take forcing to cover compelling and constraining, preventing and restraining; approving and encouraging to be covered by praising, and disapproving and discouraging by blaming; provoking to cover instigating, protesting, exhorting and urging; offering to cover inviting and bribing; advising to cover guiding and instructing; requesting to cover demanding and petitioning; persuading to cover

coaxing and cajoling; and deterring to cover hindering and impeding.

6.2. Law and Liberty

According to Bentham, "a law by which nobody is bound, a law by which nobody is coerced, a law by which nobody's liberty is curtailed, all these phrases which come to the same thing (are) so many contradictions in terms" (*L.c.*, vi, 3). This is thought to follow from his imperative theory of law (Sec. 6). According to Hart's improved model of the imperative theory, which is substantially Austinian rather than Benthamite, the form of any law is (really) (1) "Let S D ! For either S will D , or R will punish S ," where R (*Rex*) is any sovereign and S is any subject of R (Hart [1961], pp. 18ff., 234ff.). Four points in this account of a law call for comment. Firstly, on the relation of law to desire (or will). In laws, the modes in which R 's will is declared are (I) command and (II) prohibition (command to forbear) (Sec. 6). Secondly, one must speak of a general order "explicitly" backed by a threat, because the sanction which backs a penal law needs to be stated, not implied, else the law will not deter as it is intended to do (see Bentham, *L.c.*, xi, 3–4). Thirdly, one may speak of R "punishing" S for non-compliance; but not of, e.g., H "punishing" T for non-compliance. The reason for this is that in the former case the threat is executed by an *authority*, whereas in the latter case this is not so (Hart [1962], pp. 16ff.). Fourthly, it seems to me that the imperative theory could equally well represent any law as an explicit general throrder of the form (2) "Let S D ! or R will punish S " (Subsec. 6.1). This version of the theory has the advantage of being shorter and simpler than (1) (above).

However, it is false that *all* laws are of this type. Bentham's insistence that, despite appearances to the contrary, all laws are ("really," or "reducible to") imperatives, contrasts instructively with the view, mentioned above, that all imperatives are ("really," or "reducible to") indicatives (Subsec. 6.1). For instance, the Wills Act of 1837 does not *order* S to make a valid will on pain of punishment if he does not comply; rather, it *permits* and *enables* him to do so (Hart [1961], pp. 27f.). Again, general orders may be backed by offers instead of by threats. Such laws are non-coercive, since offers, being a species of permissions, do not compel (Subsec. 6.1). Bentham calls them "praemiary laws" (from the Latin *praemium*, meaning "a reward"). E.g., a law providing that any S who

reports a robbery to *R* shall receive from *R* a reward of £10. Bentham also notes that there can be "a law with alternative sanctions." E.g., a law providing that *S* shall be given a reward of £10 if he reports a robbery to *R*, but that he shall be fined £10 if he fails to do so. Manifestly, these alternative sanctions are throffers, and are of the same type as "Either I will give you my money if you will give me your hand in marriage, or I will take your money if you will not give me your hand in marriage" (Subsec. 6.1). Bentham points out, however, that orders backed by offers are much less effective than orders backed by threats. He gives several cogent and characteristic reasons why the legislator must rely far more on the stick than on the carrot, such as that "the sources of pleasure are few and soon exhausted; the sources of pain are innumerable and inexhaustible" (*l.c.*, xi, 1-10).

But even if the imperative theory does not apply to all laws, it holds approximately true of some laws. The sort of law that it fits best is a criminal statute enacted by the British Queen in Parliament. So let us take as paradigm of a law such a statute, which prohibits *S*₁ from murdering *S*₂ on pain of punishment by death (*L*). Then how, and why, does *L* restrict *S*₁'s liberty? Firstly, it does not do so *qua* order, because orders do not coerce; but it does so *qua* threat, because threats do coerce (Subsec. 6.1) (*Cp.* the quotation from Hart on the usual meaning of "unfree" in political contexts (Sec. 1).). Secondly, *L* does not diminish *S*₁'s liberty by making *S*₁ unable to murder *S*₂, but by making *S*₁ unable both to murder *S*₂ and to remain alive. (It is assumed that *L* is effectively enforced.) We have seen that Cassinelli shares the popular misconception on this point (Sec. 3). For, thirdly, the purpose of *L* is not to prevent *S*₁ from murdering *S*₂, but to *deter* him from doing so.

6.3. *Opinion and Liberty*

Mill emphasises that his "principle of individual liberty," which asserts that *C*'s only justification for coercing *A* is to prevent him from harming *B*, defines the proper bounds not only of governmental compulsion but also of the "coercion of public opinion." This raises the question whether public opinion does in fact coerce and, if so, how. The answer is that it does so because, as Mill puts it, "society can and does execute its own mandates" (*l.c.*, i). E.g., a community (*C*) makes a mandate (*M*) that any fellow (*F*) of *C* who does not attend church at least once on Sundays shall be ostracized.

M resembles *L* in the following important respects. Firstly, *M* curtails *F*'s liberty by making him unable both to absent himself from church on Sundays and also to be socially accepted. For, secondly, the purpose of *M* is not to prevent *F* from absenting himself from church on Sundays, but to deter him from doing so. On the other hand, *M* differs from *L* in the following important respects. Firstly, if *C* ostracizes *F* for non-compliance with *M*, then *C* is not rightly said to "punish" *F*, because *C* is not an authority. In this respect, *C*'s position is the same as *H*'s and different from *R*'s. Secondly, whereas the threat which backs *L* is always explicit, the threat which backs *M* is normally tacit (Subsec. 6.2). The evidence for the existence of the threat is *C*'s actions. If *C* regularly does ostracize recusants, then *F* must infer that *C* is tacitly threatening to ostracize him if he does not comply with *M*. *H*, too, can threaten *T* tacitly, i.e. by deed rather than by word. Instead of saying to *T* "Either you will give me your money, or I will take your life," *H* could silently point a pistol at *T*'s heart with one hand and extend his other hand to *T* with the palm upwards and open.

There are also praemiary social mandates corresponding to praemiary state laws. E.g., *C* makes a mandate that any *F* who attends church more than once on Sundays shall be honoured. Here, the order is backed by an offer. But, like praemiary state laws, praemiary social mandates are non-coercive because offers are non-coercive. On the other hand, my principal contention here is precisely that the concept of a *tacit threat* is a very important one, because tacit ("veiled") threats are the means whereby society normally coerces its members. With some reason, Mill considers "the tyranny of the prevailing opinion" to be a danger at least as great as "the tyranny of the magistrate" (*l.c.*).

In formulating his famous principle, Mill speaks sometimes of "preventing" *A* from harming *B*, but at other times of "detering" him from doing so. It is, of course, the latter which is correct, since it is deterrence and not prevention which is the object both of minatory state laws and also of minatory social mandates; which are what Mill is discussing. To repeat: the purpose of *L* is not to *prevent* *S*₁ from murdering *S*₂, but to *deter* *S*₁ from murdering *S*₂ by *preventing* *S*₁ from both taking *S*₂'s life and also keeping his own life. If *R*'s purpose were to prevent *S*₁ from murdering *S*₂, he could achieve it, not by making *L*, but by e.g. locking up *S*₁ and/or by locking up *S*₂ ("protective custody"). However,

the superiority of deterrence over prevention is obvious. It would be both immoral and impracticable for *R* to lock up every potential murderer, or every potential murderee, in his realm.

6.4. *Concluding Elucidations*

Some points remain in need of clarification.

The notion of a complex (conjunctive) action has figured largely in the argument of this study. However, this notion must not be confused with the more familiar idea, with which Bentham was acquainted, of alternative descriptions of an action (*l.c.*, v. 1–10). Nevertheless, the two concepts are sometimes associated. For if it is true that, e.g., *B* made *A* a tempting offer, then it is also true that “*B* made *A* an offer” and “*B* tempted *A*” are alternative (or rather, disjunctive) descriptions of what *B* did to *A*.

I say that a threat, such as *H*’s to *T*, “Either you will give me your money, or I will take your life,” curtails *T*’s liberty by making *T* about to be unable both to keep his money and also to keep his life (Sec. 3, beginning). Dr. Steiner objects that this is not so, because *H* may relent between uttering the threat and executing it; so that, generally, it is the implementing of a threat and not the making of it which diminishes freedom. I reply, firstly, that this is paradoxical, since the received view is that it is the making of threats which restricts liberty. This is particularly clear in the case of *L*, which is generally held to make *S* less free from the time of its enactment. Moreover, if *L* were to fulfil its deterrent purpose with complete success, the threat which backs it would never be executed. Consequently, on Steiner’s view, *L* would never curtail liberty; which seems highly paradoxical. Secondly, I confine the discussion to normal or genuine threats, i.e. threats which the threatener both sincerely intends and is able to execute (*l.c.*). I think that our view that the making of genuine threats abridges freedom rests on the fact that, as a general rule, threateners do execute genuine threats upon non-compliant threatenees. It is exceptional for *H* to change his mind between the utterance of his threat and the attainment of the deadline, or to forget to load his pistol. However, in a society where, as a general rule, threateners did not execute genuine threats upon non-compliant threatenees, Steiner’s view would be the received view.

Since the generalization “Almost all genuine threats are executed upon non-compliant threatenees” is statistical and not universal, one must

say in strictness of speech that *H*’s utterance of his threat *very probably* makes *T* about to be unable both to keep his money and also to keep his life. It will not do so if, e.g., *H* relents. But it is important to recognize that there is no difference in this respect between the way in which threats curtail liberty and the way in which attempted compulsion or restraint do so. For, as we have seen already, the generalization “Almost all persons who are locked in a room in *L* are unable to go from *L* to *Y*” is also statistical and not universal (Subsec. 6.1). Consequently, one must say similarly that *H*’s imprisoning *T* in *L* *very probably* makes *T* unable to go from *L* to *Y*. It will not do so if, e.g., *T* is exceptionally strong. Our concept of unfreedom, like so many others of our concepts, is relative to our idea of the average man or of human nature. Thus, we deem *A* to be unfree if he is threatened by *B*, and *B* is a normal threatener; or if *B* does to *A* the sort of thing that makes the average man retrievably unable to *D* (e.g., imprisons him, or binds him). The concepts of the average agent and of the average patient are central to the philosophy of action in the same sort of way as the notions, “the normal observer” and “normal conditions,” are central to the philosophy of perception.

This fact is connected with another important point about our concept of unfreedom, viz., that it is the concept of *some person* being made unable to perform *some sort* of action by *some other person*’s performing *some other sort* of action on him. E.g., we think of *A*’s freedom of movement being curtailed by *B*’s imprisoning him. This has a significant bearing on what Mill, Hampshire, and Steiner, e.g., call “individual liberty” (Subsec. 6.3; Hampshire [1965]; Steiner [1975]). For when we think of some individual, *H*, making some other individual, *T*, unable to perform the particular action of walking now from *L* to *Y* by performing the particular action of locking him in his bedroom now; then we think of this particular transaction as an individual instance of the general formula just stated about persons, freedom of movement and imprisonment. The concept of individual liberty is the concept of the liberty of the individual variable, *A*, and not the concept of the liberty of the particular individual, *T*. Hence, we deem *H*’s action to abridge *T*’s liberty, unless we know that *T* is strong enough to be able to break down his bedroom door.

I also say that *B*’s threat to *A* diminishes *A*’s liberty, because it makes *A* about to be unable to perform some conjunctive action which he can

unconditionally now perform (Sec. 3, end). In the case of H's threat to T, it is indisputable that, before the threat, T can unconditionally both have his money and also have his life; for he is now doing so. In the case of L, however, Professor Hart objects that it is false that, before R makes L, S₁ can unconditionally both take S₂'s life and also keep his own life (Subsec. 6.2). For, firstly, S₁ is not now killing S₂. And secondly, S₁ cannot kill S₂ unless S₁ does not die first, and unless S₂ does not kill S₁ first, and unless S₁ and S₂ are not simultaneously struck dead by lightning first, and unless . . . etc. I think that the reply to this objection is as follows. We deem A to be unconditionally able to D unless we know or believe that some preventing condition is not merely possible, but either actual or probable. Thus, it is true that S₁ cannot kill S₂ if both are simultaneously struck dead by lightning first. But since we know that this compound (conjunctive) event is *very improbable*, we do not cease to consider S₁ as unconditionally able to kill S₂ merely because we know that this compound event is possible. Consider, by contrast, the statement "Before H offers his life to T in exchange for T's money, T cannot have H's life unless H gives it to T." Here, we deem T to be only conditionally able to have H's life, because we know it is a *fact* that, unless H gives his life to T, T cannot have H's life in the relevant sense of "have H's life," viz "receive H's life from H." (See O'Connor [1971], pp. 26ff.).

Next, the subjects of liberty. It is implicit in what precedes that *what* is free is a person, and I have answered the question *when* a person is free in detail above (Sec. 5). However, we call many other sorts of things free, the most important of which are actions and wills. But I agree with Hobbes that all these uses must be explained in terms of the freedom of persons. Thus, (1) as to free action, A Ds freely if and only if A Ds and it is not the case that A is unfree to -D. (2) as to free will, A Ds of his own free will (or choice) if and only if A Ds freely; the analysis of this last formula being (1). As Hobbes says, "from the use of the word *Free-will*, no Liberty can be inferred of the will . . . , but the Liberty of the man . . ." (l.c.).

It is an important fact about liberty that there are degrees of it, as there are also degrees of potency (see Oppenheim [1961], pp. 179-210). Suppose, then, that H chains to T's left ankle an iron ball, which is not so heavy as to immobilize T completely, but is heavy enough to slow him down considerably. Does H thereby make T more unfree to walk from L to Y? No; because, firstly, the last

expression is illogical. For although we speak of A as more (or less) unfree, we do not speak of him as more (or less) unfree to D. The reason for this is, secondly, that B makes A unfree to D if and only if B makes A retrievably unable to D, and that there are no degrees of ability or possibility (Subsec. 5.5). However, this last fact is liable to be obscured by improper manners of speaking. E.g., a sufferer from arthritis is apt to say "I am not so able to move about as I was" instead of "I am not able to move about so much as I was." Nor is this error exclusive to the vulgar. Philosophers of probability, for instance, are prone to speak of "equally possible" alternatives. Consequently, A becomes more (or less) free, or more (or less) powerful, by becoming able to do more (or less) things; not by becoming more (or less) able to do any given number of things (Sec. 4, Subsec. 5.5). What H achieves by means of the ball and chain is to *impede* T's walking from L to Y. Hobbes' definition of freedom is open to criticism on this score, since he speaks of A being free to D if he is not "hindered" from D by B (l.c.). But "hindered" means the same as "impeded," and Hobbes needs for his definition a stronger verb, such as "prevented" or (as he himself says elsewhere) "stopped" (cf. Subsec. 5.1, end).

Finally, a word on the importance of freedom. Oppenheim remarks that ". . . 'liberty' is perhaps the word which occurs most frequently in political life and in political theory" (Oppenheim [1961], p. 109). Why is this? Part of the answer, without doubt, is the high value that men set on liberty. In this essay, however, I have virtually ignored the evaluative aspect of the philosophical problem of liberty. For instance, we have considered whether laws do curtail liberty; and if so, how exactly, and why. But nothing has been said about the great and complex question of whether laws ought to be used to curtail certain sorts of liberties; and if so, to what extent, and how exactly, and why. I deliberately say nothing here about the *right* to liberty, because it is quite task enough for one study to explain the *meaning* of "free." Nevertheless, it is relevant to add that the very fact that liberty is a right throws into relief what we have seen to be an essential feature of freedom, viz., its interpersonal character. For a right is always held or claimed by some person, A, against some other person, B. Or, to put the same point differently, it is a telling objection against Plato's intrapersonal, master/slave conception of unfreedom that it carries with it the absurd consequence that one can speak meaning-

fully of one part of *A*'s soul invading the right to liberty of another part of it (Subsec. 5.3).

We have seen that Plato's view rests on metaphor, and there is another common metaphor which deserves notice. *A* may tell *B* that he is unfree to dine with him tomorrow because he has engaged himself to dine with *C*. This is literally false, because *A* can break his engagement to *C* and dine with *B* instead. Suppose, however, that *R*

makes a law which provides that, if *S*₁ breaks his engagement to marry *S*₂, then he must pay *S*₂ damages. Then *S*₁'s liberty is curtailed, because he cannot both break his engagement and also keep (all of) his money. But what curtails *S*₁'s liberty is not his promise, but *R*'s law. Talk of "binding promises" and "moral laws" is as-if talk, like talk of "slavery to the passions" and "irresistible offers."

Keele University

Received February 27, 1976

REFERENCES

References are given in the text between brackets, either (a) in the case of classic works, by part, chapter, section etc.; or (b) in the case of other works, by page number to the works listed below. No works published after 1975 are included in this list.

- AUSTIN, J. L., "Ifs and Cans," *British Academy Proceedings*, vol. 42 (1956), pp. 109-132.
 ——— *How to do Things with Words* (Oxford 1962).
 AYER, A. J., "Freedom and Necessity," *Polemic*, no. 5 (1946), pp. 36-44.
 BAY, CHRISTIAN, *The Structure of Freedom* (Stanford, 1958).
 BENN, S. I., "Freedom and Persuasion," *The Australasian Journal of Philosophy*, vol. 45 (1967), pp. 259-275.
 BENN, S. I. and PETERS, R. S., *Social Principles and the Democratic State* (London, 1959) chs. 9, 10.
 BENN, S. I. and WEINSTEIN, W. L., "Being Free to Act, and Being a Free Man," *Mind*, vol. 80 (1971), pp. 194-211.
 ——— "Freedom as the Non-Restriction of Options," *Mind*, vol. 83 (1974), pp. 435-438.
 BERLIN, ISAIAH, "From Hope and Fear Set Free," *Aristotelian Society Proceedings*, vol. 64 (1964), pp. 1-30.
 ——— *Four Essays on Liberty* (Oxford, 1969).
 BOSANQUET, BERNARD, *The Philosophical Theory of the State* (London, 1923).
 BRADLEY, F. H., "My Station and its Duties" in *Ethical Studies* (Oxford, 1927).
 BROADIE, ALEXANDER, "Imperatives," *Mind*, vol. 81 (1972), pp. 179-190.
 CASSINELLI, C. W., *Free Activities and Interpersonal Relations* (The Hague, 1966).
 CRANSTON, MAURICE, *Freedom: A New Analysis* (London, 1953).
 CRICK, BERNARD, "Freedom as Politics" in *Philosophy, Politics and Society*, series 3, ed. by Peter Laslett and W. G. Runciman (Oxford, 1967).
 DAY, J. P., "Unconscious Perception," *Aristotelian Society Supplement*, vol. 34 (1960), pp. 47-66.
 ——— "On Liberty and the Real Will," *Philosophy*, vol. 45 (1970), pp. 177-192.
 DUNCAN-JONES, AUSTIN, "Freedom to do Otherwise," *The Cambridge Journal*, vol. 3 (1950), p. 750.
 DWORKIN, GERALD, "Acting Freely," *Noûs*, vol. 4 (1970), pp. 367-383.
 FAIR, HASKELL, "Prediction and Constraint," *Mind*, vol. 67 (1958), pp. 366-378.
 FEINBERG, JOEL, *Social Philosophy* (Englewood Cliffs, 1973), chs. 1-3.
 FRANKFURT, H. G., "Coercion and Moral Responsibility" in *Essays on Freedom of Action*, ed. by Ted Honderich (London, 1973).
 ——— "Three Concepts of Free Action," *Aristotelian Society Supplement*, vol. 49 (1975), pp. 113-125.
 FRIEDMAN, R. B., "A New Exploration of Mill's Essay On Liberty," *Political Studies*, vol. 14 (1966), pp. 281-304.
 GREEN, T. H., "On the different senses of 'Freedom' . . . etc." in *Lectures on the Principles of Political Obligation* (London, 1937).
 HAMPSHIRE, S. N., *Freedom of the Individual* (London, 1965).
 HARE, R. M., *The Language of Morals* (Oxford, 1952).
 HART, H. L. A., "Prolegomenon to the Principles of Punishment," *Aristotelian Society Proceedings*, vol. 60 (1960), pp. 1-26.
 ——— *The Concept of Law* (Oxford, 1961).
 ——— *Law, Liberty and Morality* (London, 1963).
 ——— "Bentham's *Of Laws in General*," *Rechtstheorie*, vol. 2 (1971), pp. 55-66.
 HART, H. L. A. and HONORÉ, A. M., *Causation in the Law* (Oxford, 1959).
 LOCKE, DON, "Three Concepts of Free Action," *Aristotelian Society Supplement*, vol. 49 (1975), pp. 95-112.

- LYONS, DANIEL, "Welcome Threats and Coercive Offers," *Philosophy*, vol. 50 (1975), pp. 425-436.
- MABBOTT, J. D., *The State and the Citizen* (London, 1947) pt. B.
- MACCALLUM, G. C., "Negative and Positive Freedom," *The Philosophical Review*, vol. 76 (1967), pp. 312-334.
- MACMURRAY, JOHN, *Conditions of Freedom* (Toronto, 1949).
- MCCLOSKEY, H. J., "A Critique of the Ideals of Liberty," *Mind*, vol. 74 (1965), pp. 483-508.
- MCCLAUGHLIN, R. N., "Deontic Logic and Conditional Obligation," *Mind*, vol. 82 (1973), pp. 207-217.
- MELDEN, A. I., *Free Action* (New York, 1961).
- MOORE, G. E., *Ethics* (London, 1912), ch. 6.
- NOWELL-SMITH, P. H., *Ethics* (London, 1954), chs. 19, 20.
- NOZICK, ROBERT, "Coercion" in *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*, ed. by S. Morgenbesser et al. (New York, 1959).
- O'CONNOR, D. J., *Free Will* (London, 1971).
- OPPENHEIM, F. E., *Dimensions of Freedom* (New York, 1961).
- PARENT, W. A., "Freedom as the Non-Restriction of Options," *Mind*, vol. 83 (1974), pp. 432-434.
- "Some Recent Work on the Concept of Liberty," *American Philosophical Quarterly*, vol. 11 (1974), pp. 149-167.
- PENNOCK, J. R., "Coercion: An Overview," *Nomos*, vol. 14 (1972), pp. 1-15.
- PLAMENATZ, J. P., *Consent, Freedom and Political Obligation* (Oxford, 1938), chs. 5, 6.
- POLLOCK, LANSING, "Freedom and Universalizability," *Mind*, vol. 82 (1973), pp. 234-248.
- RYLE, GILBERT, *The Concept of Mind* (London, 1949), ch. 3.
- SKINNER, B. F., *Beyond Freedom and Dignity* (New York, 1971).
- STEINER, HILLEL, "Individual Liberty," *Aristotelian Society Proceedings*, vol. 75 (1975), pp. 33-50.
- WEINSTEIN, W. L., "The Concept of Liberty in Nineteenth Century English Political Thought," *Political Studies*, vol. 13 (1965), pp. 145-162.
- WHITE, D. M., "Negative Liberty," *Ethics*, vol. 80 (1970), pp. 185-204.
- WILSON, JOHN, "Freedom and Compulsion," *Mind*, vol. 67 (1958), pp. 60-69.

ACKNOWLEDGMENTS

This essay is dedicated to Professor Herbert Hart, to whom I am grateful for criticizing an earlier version of it, and for many other acts of kindness. I am also grateful to Professors Gilbert Ryle and Antony Flew, and Dr. Hillel Steiner for their comments on it. Earlier versions were discussed at a Joint Seminar of the Departments of Law, Philosophy and Politics of Keele University on 29 April, 1975; and at a Senior Political Theory Seminar of Manchester University on 13 February, 1976. I am obliged as well to participants in those seminars for their suggestions.

II. PERCEPTION: A NEW THEORY

EDMOND WRIGHT

THE Philosophy of Perception has been bedevilled by two opposing faults. On the one hand its practitioners have often not set themselves the aim of a true generality: they have too often allowed themselves unthinkingly to be bound to the facts of the human condition; they have failed to discuss perception as it could logically apply to *any conscious being anywhere*. On the other hand, they have failed in not being empirical enough: they have tended to judge perception merely from their own personal experience, ignoring obvious empirical truths.

An example of the first would be the tendency to judge all perception from the stance of an individual who, as a result of his socialization-past, is now reasonably mature and stable in his society, applying its concept- and percept-rules without any serious qualms; Austin's stress on the standard-type¹ could be said to be of this category. An example of the second would be the confident rejection of any compatibility between mental-imaging and the sensing of the external world, to the point, as notoriously with Ryle, of reducing mental-imaging to a mere pretending that we are seeing or hearing or touching and so on.²

In order to arrive at a proper philosophical stance we must strive at once to be both *general*, in not confining ourselves to the experience of a mature human being, and more *empirical*, in taking into account reports from other people whose sensory experience might in fact be empirically impossible for ourselves.

The new theory to be presented here attempts to reconcile opposing views: to satisfy the realist who wishes to preserve the "primary objects of perception,"³ the materialist who wishes to find a place for the "scientist's eye,"⁴ and for the phenomenologist or the sense datist with their championship of the subjective "raw-feel" experience.⁵ It is an application of *irony* which has already been used of meaning and truth.⁶ The word "irony" with regard to perception was first used by Sam Coval and D. D. Todd in 1972,⁷ and the argument here is indebted to their insight; also to that of Fred I. Dretske's discussion of radical relativity in perception,⁸ to a recent paper by J. B. Maund,⁹ to G. J. Warnock's "Seeing" and the postscript thereto,¹⁰ and to an insistent query by G. N. A. Vesey, whose own sentence "All seeing is seeing as" will be given a special interpretation here. The nature of irony as it appears in situations and verbal exchanges is to be used as a model—twice over—for what happens in perception. This model, it is claimed, will answer Vesey's own demand that "there simply has to be a 'looking' and a 'seeing' which has a foot in both camps [epistemic and non-epistemic] but not both feet in either, if seeing is to be a reason for believing."¹¹ It is the hope that the theory presented here does precisely that, and with the pattern of the duck-rabbit that Vesey himself suggests.

Before beginning the main argument, I have to tilt at two rickety windmills. Since philosophers have been accused of lacking imagination, let us

¹ J. L. Austin, *Sense and Sensibilia*, ed. by G. J. Warnock (Oxford, 1962).

² Gilbert Ryle, *The Concept of Mind* (London, 1949), Ch. VIII.

³ David H. Sanford, "The Primary Objects of Perception," *Mind*, vol. 85 (1976), pp. 189-208.

⁴ D. M. Armstrong, *A Materialist Theory of Mind* (Boston, 1968).

⁵ Jonathan Harrison, "Direct Perception and the Sense-Datum Theory" in *Contemporary British Philosophy: Fourth Series*, ed. by H. D. Lewis (London, 1976), pp. 132-151.

⁶ Edmond L. Wright, "Arbitrariness and Motivation: A New Theory," *Foundations of Language* (forthcoming, 1976); "Illusion and Truth," *Philosophy and Phenomenological Research* (forthcoming, 1977).

⁷ Sam Coval and D. D. Todd, "Adjusters and Sense-Data," *American Philosophical Quarterly*, vol. 9 (1972), pp. 107-112.

⁸ Fred I. Dretske, *Seeing and Knowing* (London, 1969), Ch. V.

⁹ J. B. Maund, "The Non-Sensuous Epistemic Account of Perception," *American Philosophical Quarterly*, vol. 13 (1976), pp. 57-62.

¹⁰ G. J. Warnock, "Seeing" in *Perceiving, Sensing and Knowing*, ed. by Robert J. Swartz (Garden City, 1965), pp. 49-67.

¹¹ G. N. A. Vesey, "Analysing Seeing (II)" in *Perception: A Philosophical Symposium*, ed. by F. N. Sibley (London, 1971), pp. 133-137 (see p. 137); see also "Seeing and Seeing As" in Swartz, *op. cit.*, pp. 68-83.

start with a little excursus into imaging itself. Stewart Candlish¹² has already argued convincingly that mental imaging and ordinary sight are not incompatible, and O. R. Jones¹³ that after-images are similar in sensory character to ordinary sight. It is not surprising therefore that a good visile like myself is able to perform the following.

I look at the sun for a while and then close my eyes, still keeping them facing in the direction of the sun. What do I see? The after-image of the sun, a small livid patch of a purplish tinge, set against a dull glowing unevenly red background, this latter being the sunlight viewed through the flesh of my eyelids. I now set my mental imaging to work: I see a small boat sail round the purple patch, as if the purple patch is an island; a sailor on board looks up at the red "clouds" and scratches his head—I am making him wonder whether it is going to rain. Of course, as any good visile will tell you, I see him in full colour, somewhat spoiled in this case through being imaged over a dull red background, but coloured nevertheless—there is his *blue* sail, for example. Here, in spite of those philosophers of perception listed by Candlish, I am experiencing 1) mental imagery in reasonably good colour; 2) an after-image (here played with as the island round which the sailor is going in his boat); and 3) real sight, namely the dull red of the sunlight seen through flesh (here taken as the clouds). So I am seeing all three simultaneously. Those who have argued a *qualitative* difference between these three are plainly wrong in empirical fact.

A simpler, less extraordinary example would be an experience of mine two years ago while I was typing out my thesis. I would wake in the morning, look up at the ceiling rose of the lamp to see it clearly, though faintly, surrounded by typewriter keys. *Both* the ceiling rose *and* the typewriter keys were visible together. This was not an example of a freely willed mental imaging, so there are exceptions to one of the conditions Berkeley depended on.¹⁴

A second windmill on the point of collapse, if it has not already done so, is the belief that mental images cannot be inspected. There are several statements of this to be found, but I will quote one only somewhat forcibly put: "We all believe *esse* is

percipi in the case of mental images."¹⁵ This implies that we can only image a *perception*—hence we cannot inspect it, for we already have! But here is a practical disproof.

When in my teens I became a fervent player of classical music on gramophone records, I very soon found that I could play through in my head the whole of pieces that I was particularly fond of and had heard countless times. For example, without any difficulty I came to be able to hear the whole of Sibelius's tone poem *En Saga* with complete orchestration, and still can. In listening to this I am clearly not merely "following a tune," as Ryle puts it:¹⁶ I am experiencing the work as complete in tonal range and volume variation as on a gramophone record. For Ryle to say that "a deafish man may follow a tune better than one who hears it better" is mistaken.

If I were a deafish man—let us say that my ears were such that there was a cut-off at 4 kilo-hertz—I could just not hear the orchestration as well as I do. Now I am not a student of music. Let us suppose that I became one, and in the course of my studies I came to understand what a plagal cadence was. Without any further recourse to a score or a performance or a gramophone record of *En Saga*, I now proceed to play it over in my head, and in the course of my doing so now write down all the plagal cadences I come across. Perhaps some of them are exceedingly difficult to discern, because they depend on the hearing of a piccolo note gliding together with one from a tuba while a glockenspiel is playing forte in the background. Now if Ryle's analysis of hearing a tune in one's head were no more than "using my knowledge of how a tune goes"—an impenetrable series of "abstentions" from producing the notes—this re-interpretation and inspection of my auditory image would be impossible. The deafish man would not even hear the high notes played on the piccolo and the glockenspiel, that is, he would not have been able to hear them in an original performance or from the playing of the gramophone record, and obviously could not have them as part of his mental image (we are assuming that the deafish man has had this 4 kilo-hertz cut-off since birth). There would in fact be nothing that a music teacher could

¹² Stewart Candlish, "The Incompatibility of Perception: A Contemporary Orthodoxy," *American Philosophical Quarterly*, vol. 13 (1976), pp. 63–68.

¹³ O. R. Jones, "After-images," *American Philosophical Quarterly*, vol. 9 (1972), pp. 150–158.

¹⁴ George Berkeley, *The Principles of Human Knowledge*, Sections 29 and 30.

¹⁵ C. W. K. Mundle, *Perception: Facts and Theories* (London, 1971), p. 88.

¹⁶ Ryle, *op. cit.*

ask me to do with *En Saga* that I could not do *as well* with the mental image as with a record: indeed, it would be better done with the mental image because it is entirely fluid under the control of my will—parts of it could be left out, different orchestrations tried, alternative modulations, tempi, and so on. The conclusion is clear: mental images are inspectable. The argument is easily generalizable to sight: I learn to mental-image a page of Chinese writing; someone teaches me Chinese; I re-activate the mental image and read the writing.

If these are facts, then an improvement on the experiment suggested by Jonathan Harrison should be possible.¹⁷ Harrison describes a possible experiment in which a subject is shown pictures of lions on a screen, and then, unbeknownst to the subject, the screen is rendered transparent to show real lions behind. The subject, he rightly claims, would have difficulty in distinguishing the two. Let us go further, knowing what we now know about compatibility and inspectability. Probably no human agent could perform this, but it is logically possible that some agent could.

There are three screens side by side in a darkened auditorium. We show an expert visile, before he goes in, a red billiard ball lying in a well-lighted cubical box lined with green baize. He views it from the side, as if he were looking into an aquarium. He sees it long enough to learn to be able to cast a vivid image of it for himself onto Screen No. 1. He is asked to make the size of the image of the open side of the box exactly match the size of the screen. We tell him to hold the image in full colour stereopsis over Screen No. 1, and our expert visile does so. He is so expert a visile that he can even place a mental image anywhere in his visual field, right in the centre of the fovea or anywhere on either side (remember we are dealing with logical possibilities). In fact to him Screen No. 1 now looks no different from the box itself when he saw it.

Our experimenter now throws a 3-D colour picture of the very same box onto Screen No. 2 (in our imagined laboratory they can do this without the onlooker having to wear special glasses). Screen No. 3 now becomes transparent to reveal the actual box containing the original red billiard ball. All three presentations look *exactly the same*.

Take another logical possibility. Suppose our experimenter had such neurological knowledge that, without hurting the subject and without his knowing what was happening, he was able to displace the subject's mental image *from one screen to another* instantaneously. The experimenter can now interchange 3-D picture, mental image, and reality from screen to screen (all screens being such that they could be rendered transparent without the subject begin able to tell. The subject could only tell which was the mental image by ceasing to image, when one of the screens would go blank, but he would not know which one was going to!

Allow our logically possible experimenter the extra neurological knowledge by which he could prevent the subject cancelling his mental image (by the subject having been secretly given a harmless short-acting drug beforehand), and the subject would have no idea which screen showed the real billiard ball, which the 3-D picture, and which the mental image in full colour stereopsis. To simplify matters for those still worried about possibilities here, the subject is not allowed to move and, by some harmless means, his eye movements are limited to viewing those three screens only from a constant position.

This, of course, leads to a disturbing conclusion. We are given the Cartesian problem of the reality of the external world in a peculiarly vivid form. We can no longer rely on the "vivacity," "steadiness," or "controllability by will" of the impressions concerned. Are we at the sceptic's mercy? Are you a solipsist?

It is plain that there is something in common to all three presentations. Coval and Todd are right to claim that there is something "neutral" between illusion and open-eye sight "to enable the very comparison to be made."¹⁸ This neutral something is *non-epistemic seeing*. Roderick Firth was right when he argued against Austin about Macbeth's dagger. He took Austin's challenging question—"Can the question seriously be raised whether the dagger which is said to be seen is bodily, or just looks bloody?"—and answered, "but there could be no difference between these questions when Macbeth *does not know* whether he is seeing a dagger or an hallucination."¹⁹ Our subject in the experiment does not know whether the screen is showing something that *looks like* a billiard ball (3-D pic-

¹⁷ Harrison, "Direct Perception," *op. cit.*, p. 150.

¹⁸ Coval and Todd, "Adjusters," *op. cit.*, p. 112.

¹⁹ Roderick Firth, "Austin and the Argument from Illusion," *The Philosophical Review*, vol. 72 (1964), pp. 372-332 (see p. 381).

ture), is an hallucination (the artificially fixed mental image), or is the billiard ball. What proof have we then of the external world?

Let us now move to the central duck-rabbit model. As I have already used the duck-rabbit elsewhere, let us take another ironic picture to effect the same analogy. Here is a mark of sorts:

Figure 1

We now place another mark beside it:

Figure 2

Clearly, a naive and unsuspecting subject would tell us, if shown Figure 2, that he sees the number *sixty-one*. Let us show some other unwary subject the following:

Figure 3

He will tell us that he reads the word "is." Suppose to the third group we now show this:

Figure 4

They will complain, and rightly. Is this "sixty-one ess" or "six-IS"?—or even "six-eye-ess." The central mark is a duck-rabbit.

We can analyse Figure 4 into *five* elements as follows:

1. the *sensory presentation* itself (known by us as "the original upright mark," but—and this is important, as will be understood later—it might have had *no* clear interpretation as a "mark" or "line" or anything else);

- 2(a). a *context-indicator A* (a numeral, the '6')—which leads us to perceive

- 2(b). *interpretation A* of No. 1, namely the numeral '1' of the number "61";

- 3(a). a *context-indicator B* (the letter 'S')—which leads us to perceive

- 3(b). *interpretation B* of No. 1, namely the letter 'I' of the word "IS."

To keep the model simple I am ignoring the other possibilities (such as "six-eye-ess" or "six-one-ess" or any other). Now exactly the same pattern of analysis can be applied to any *irony* of this basic dual kind. Take a pun:

"He walked with a pronounced limp—
L-I-M-P—pronounced 'limp'."

(Spike Milligan)

We analyse as follows:

1. the *sensory presentation*: /prounaunst limp/ (It has been given in phonetic symbols in order to indicate that no interpretation is being implied in our analysis here; notice also that it occurs twice in the statement, but the ironic effect plays over both occurrences as soon as context-indicator B appears);

- 2(a). *context-indicator A*: "He walked with a . . ."—which leads us to perceive

- 2(b). *interpretation A*: "a marked unevenness of gait, caused by some injury or defect" (this awkward paraphrase is only an attempt to specify without ambiguity the first meaning taken by the hearer);

- 3(a). *context-indicator B*: "L-I-M-P" (that is, the letters spelled out separately), leading us to perceive

- 3(b). *interpretation B*: "uttered in this manner—'limp'."

You might try a similar analysis for yourself on these varied examples:

- (i) "I do most of my work sitting down: that's where I shine."

(Robert Benchley)

- (ii) (From a tailor's advertisement) "Why not hire-purchase one of these best-quality suits you've so rightly admired? When you buy one of our suits, you deserve a little credit!"
- (iii) IAGO: . . . Proclaim him in the streets; incense her kinsmen;/And though he in a fertile climate dwell,/Plague him with flies.

Now let us carry our analysis of irony into the Philosophy of Perception. For simplicity's sake we

will keep to the basic dual irony with its ambiguous sensory element, its two-context-indicators, and its two interpretations—a five-fold analysis. How does this Irony Model fit perception?

First, it applies to our *sensing* itself, our *non-epistemic* seeing. Harrison²⁰ is the most recent writer on this topic to stress the subjective differences between one observer and another. What has to be accepted is that each of us makes a different bodily *registering* of the sensation-producing forces that we have systems to detect. It is not only a matter of perspectival difference, though that is part of it: the actual sensory registering is going on in different bodies with their own empirical peculiarities. One simple, but I think conclusive demonstration of this empirical fact I can exemplify with an experience common to myself and my son. We both have a curious genetic characteristic affecting our sight: our right eyes are more sensitive to blue light; our left eyes are more sensitive to red light. If I cover up my right eye and look with my left, I see a *different red* from when I cover up my left eye and look with my right. My son reports the same effect. If my two eyes can be different from each other in "raw-feel" quality, how much more likely is it that two persons differ!

There is a further point to get clear here. Even if a psychologist were able to show that in band width, spectrum distribution and intensity response two people were so measurably close that any difference was beyond their "just-noticeable-difference" level, we must still say that there are two numerically different registrations. That supposedly suspect TV analogy is quite appropriate here (and elsewhere as we shall see): two TV cameras may take two exactly similar shots of the same scene and on their closed circuit screens the same picture may show, yet still it makes sense to say that there are two pictures produced by different electrons on two different screens. Compare Renford Bambrough's story of the child faced with a strip cartoon (a teddy-bear's adventures) for the first time:²¹ "Here's a teddy-bear!" says the child of the first picture, and "Here's another teddy-bear!" of the second; we hasten to correct the child—"It's the *same* teddy-bear!" we say—just as we do of *any sensory view* we share.

So the Irony Model must be applied to sensing.

Of two persons both sensorily affected by sensation-producing forces within their overlapping views, there is

1. a *region of matter* by which they are both sensorily affected (and here we are leaving aside for the time being the all-important problem of solipsism and the reality of the external world—the question that is begged here will be *answered*);

2(a). a *physiological sense-receptor system A* embodied in a material agent with its own peculiarities (e.g. in distinctness of focus, sensitivity to change and intensity, band width, etc.), giving

2(b). a *private subjective registering A* within that material agent.

3(a). another *physiological sense-receptor system B* embodied in another material agent with . . . etc., giving

3(b). another *separate subjective registering B* within that second material agent.

So these are empirical facts that cannot be ignored. There is a systematic ambiguity when any two agents claim that they are sensing something in common. They agree, to use Alfred Schutz's phrase,²² to "idealize their reciprocity," to act as if their subjective registrations were absolutely the same. Just as it is vital for the story in the strip-cartoon that we take the next teddy-bear as the *same* teddy-bear, it is so often vital for us to assume that we are sensing in exactly the same way, and so often it happens that any difference can be ignored. This is the case when the purposes of the two agents as evidenced in preceding behaviour have been shown to be in apparent agreement, and the difference in subjective registrations has so far not been found to be relevant to those purposes. But the logical possibility remains open that the difference in subjective registrations might be significant for behavior.

An important point about this application of the Irony Model is that 2(b) and 3(b) are not words or meanings, but "raw-feel" *experiences*. Those like Ayer, who have tried to find a sense-datum *language* are in error. Benson Mates²³ is correct in saying that what we are applying our percepts to is not some parallel sense-datum *language* of "specks" or "lumps" or "crescent-shapes" or what-have-you, but to sensory experience itself. His using the old

²⁰ Harrison, *op. cit.*, pp. 148–149.

²¹ Renford Bambrough, "Appearance, Identity and Ontology," *Proceedings of the Aristotelian Society*, vol. 75 (1975), pp. 69–76 (see p. 73).

²² Alfred Schutz, *Collected Papers, Vol. I, The Problem of Social Reality* (The Hague, 1962), pp. 31–32.

²³ Benson Mates, "Sense-Data," *Inquiry*, vol. 10 (1967), pp. 225–244.

word "denotation" does not wholly clarify matters, but at least he is on the right lines. Anyone who has to refer to the variations of the sensory field might have to use terms like "speck" or "lump", but these are only word-tools to help the *philosopher* in his analysis. The specks and lumps as experienced remain experiences not necessarily already pre-sorted *for the agent* into any such divisions. Variations, describable in infinite ways, might exist within a sensory field, but yet the distribution is in one sense at any one instant fixed and unchangeable. Price was right to insist upon this,²⁴ but it has not been easy to understand.

It might help the understanding of this point if I explained why it is possible to have an *accurate* photograph of a *blur*. (This argument owes much to Alistair Hannay's article, "To See a Mental Image."²⁵) Suppose we take a photograph of a chess-board, but we deliberately disturb the lens so that what is thrown upon the photographic film is a blurred image of the chess-board. We develop this photograph—let us say we produce a print of the same size as the original chess-board. It is now a pattern consisting of what we might call, if we felt we had to, a set of blurred black and white squares. It can also be called, if you should wish *for any purpose*, a blurred photograph of a chess-board.

We now give it to an expert in chromato-physics who subjects it to a searching series of measurements such that he has a precise record in quantified form (and, an important point, so accurate that it goes beyond the "just-noticeable-differences" that human eyes are capable of detecting) of the distribution of shade across the picture, particularly at the blurred edges of the "squares."

We now put *this picture* where the original chess-board was and photograph the picture itself, only this time we *do not put the lens out of focus*. We get an expert photographer to take a photograph which is *perfectly focussed* on the picture of the blurred squares. We now develop this second photograph and produce a print of the same size as the first photograph. Call the first 'A' (the badly focussed picture of the chess-board itself); call the second 'B' (the perfectly focussed picture of A). We take the second photograph, B, to our expert in chroma-

to-physics, who subjects it to the same tests he used on A. Our photographer (of B) is pronounced to be extraordinarily good! The measurements made on the distribution of shade across B, particularly at the edges of the "squares," exactly correspond to those across A. So exact is the reproduction that the chromato-physicist, though *he* can detect the difference between A and B with his delicate instruments, pronounces it far beyond the range of the "just-noticeable-differences" of the human eye. We have now a *perfectly focussed image of a blur*.

The analogy with mental imagery and open-eye sensing has probably not escaped you. Both are "perfectly focussed" in this second sense because they could never be otherwise. Herein lies their basic similarity. If I either mental-image or open-eye-see what I might call a "blurred patch," it is not blurred in any epistemic sense until I try to judge it as a blurred *perception* of something—for, supposing it to be a "blurred white patch," I might easily take it as a *perfectly focussed* image of a cloud of steam! As a picture of a snowball it would be blurred: as a picture of a cloud of steam it is perfectly in focus.

So what does not change from the non-epistemic point of view is the visual field. This is why the Law of Excluded Middle does not apply to it: what is a clear view of a cloud of steam can change into that of a blurred snowball *without alteration*. It was the LEM's not applying to it that led to Winston G. F. Barnes's puzzlement,²⁶ from which he incorrectly concluded that the non-epistemic field does not exist. But it just *is*, in Price's sense;²⁷ its distribution at any one instant is a fixed and unchangeable field *from which* perceptions can be drawn.

One further objection about this *first* application of the Irony Model can be answered by being determinedly general, breaking away from the empirical human limits. "What are the logical limits of such an ironic sensory overlap upon a region of matter?" The answer is clearly the case in which two agents have *entirely different* sensory systems. A being who could only hear, feel, and taste, would still be able to interact meaningfully with a being who could only see, smell, and experience kinesthetic sensation. W. J. Thomas²⁸ has

²⁴ H. H. Price, *Perception* (London, 1932), pp. 148–150.

²⁵ Alistair Hannay, "To See a Mental Image," *Mind*, vol. 82 (1973), pp. 161–182.

²⁶ Winston G. F. Barnes, "The Myth of Sense-Data," *Proceedings of the Aristotelian Society*, vol. 45 (1944–45), pp. 89–117.

²⁷ Price, *op. cit.*

²⁸ W. J. Thomas, "Communication without Sensory Overlap," *The Journal of Philosophy*, vol. 72 (1975).

already argued convincingly that we would be able to communicate with beings from another planet who had different sensory systems—because we would be able to *gesture*. Our physical movement of matter, *seen* and *felt* by us, would be ϕ 'ed and ψ 'ed by the other beings. Mundle's argument for a special status for the body-image is therefore mistaken: if other people can check a blind man's limb position by sight as he can by kinesthetic sensation, there is an overlap of the kind that will fit the Irony Model.

But the Irony Model does not apply only to sensing, non-epistemic experience: it applies to perception, epistemic experience. The "61S" example itself was a demonstration of it, but let us take one that does not involve interpreting a *sign*, because that introduced an extra, and for our present purposes, an unnecessary level of complexity.

Suppose an entomologist, an expert on moths, decides to take me out on a field trip. He takes me to a special site that he knows about, a copse of particular trees, looks about for a while, and then draws my attention to the bark of one of the trees; it is covered with patches of moss and lichen. He asks me what I *see*. Now this question in this context does not mean "Describe in 'fuzzy-lump' language the distribution of your stereoscopic visual field"; he is not asking me to describe in some object-neutral language my non-epistemic seeing. He is asking me to describe my epistemic seeing, my perceptual judgments and my material beliefs. So I say, looking at the small area that he is indicating, "I *perceive* some roughnesses of bark, some patches of moss, some platelets of lichen, some small spots of orange fungi, and some shining patches of damp." All these are very irregular in shape. He asks me to make a detailed coloured picture of it. Assume that I am an expert artist and I do so. He now puts his finger lightly on the bark and a moth takes wing. He catches it and shows me the subtlety of the camouflage. It was indeed *wholly* in my visual field all the time; nothing was intervening between myself and it. He is now even able to show me its very outline on the careful *trompe d'oeil* picture I made. He discusses patch variation on such moths, shows me a book with illustrations of the characteristic forms of such variations, and soon I can be on the perceptual watch for such moths as well as he. This example is

similar to that of the deer given by J. F. Soltis,²⁹ who used it to sustain an undefined account of what he calls "simple seeing" as always discriminatory of *objects*.

F. N. Sibley³⁰ brings out the difficulties for the epistemic theorist in a case like this, but without taking up a non-epistemic position. Part of the reason for his tentativeness perhaps lies in his still talking of the *object* (in the following quotation referred to as 'D'—the equivalent of the moth) being seen by the uninstructed observer (Soltis's hunter—'S'):

it seems to me very odd to insist that S does not see D; it is almost as if, although S does all that is needed to qualify as seeing D, one must not *say* that he sees it. If we do not call this relationship of a person to an object *seeing*, it is hard to know either what could be called seeing or what this relationship is.

David H. Sanford,³¹ commenting on Sibley, is no nearer a solution, for he makes the same mistake in believing that D is perceived as an *object*. He says that the deer was partially concealed (it is not clear in Soltis whether this is in fact the case), but he seizes upon this, claiming that, because S only sees the *flank* of the deer, it is merely a matter of his failing to apply a "perceptual rule." Sanford says that S certainly sees the flank of the deer.

Both Sibley and Sanford are confused. Of the moth example, ask this question: "Who can *say* that the moth is being seen by Wright (in the period before I was told about its presence)?" Only the entomologist. Afterwards, I myself can inspect my picture (or my vivid mental image) and pick out the new gestalt, but I cannot say that I *perceived* the moth, epistemically-saw it, in the uninstructed period. It does violence to thought to say that in that period I was *perceiving* the moth. It would be like saying that I saw the point of Spike Milligan's pun before he got to "L-I-M-P." My eyes were registering the non-epistemic seeing that included it, but I had not sorted those variations into the percept Moth.

Daryl Close³² has tried to argue that I would have a belief about the *moth*, but this is quite ludicrous. I may have perceptual beliefs about moss, lichen, fungi, bark wet and dry in careless intermingled profusion, but no one could claim that any of these beliefs *included* the moth *for me*. Close's argument (about an aunt seen as a manne-

²⁹ J. F. Soltis, *Seeing, Knowing and Believing* (London, 1966), pp. 33ff.

³⁰ F. N. Sibley, "Analysing Seeing (I)" in Sibley, *op. cit.*, pp. 81-132 (see p. 112).

³¹ Sanford, *op. cit.*

³² Daryl Close, "What is Non-Epistemic Seeing?", *Mind*, vol. 85 (1976), pp. 161-170 (see p. 168).

quin) allows him to try to make *false belief* a necessary condition of such seeing, that an epistemic element is essential:

since S takes his aunt to be a mannequin, he certainly believes of his aunt that she is a mannequin (S believes that what he sees is a mannequin); hence, this is not a case of seeing in which S has no visual beliefs at all about what he sees.

But applied to the moth case—and it should be so generalizable—the argument breaks down. *I* was not believing of the moth that it was roughly a quarter moss, half dry bark, an eighth damp bark, a sixteenth orange fungi, and a sixteenth lichen. *I* had no visual beliefs at all about the moth! How could *I* have since *I* didn't know it was there? The entomologist knew, for he had already sorted his non-epistemic seeing according to a familiar gestalt. He could say all that Sibley, Sanford and Close want to say, but only with this interpretation: "Wright was seeing the moth non-epistemically as part of a visual field which Wright was perceptually organizing into quite a different interpretation. There was no correspondence between my epistemic perceiving of the moth and his epistemic perceiving of the bark, moss, lichen and fungi except in the structure of our non-epistemic seeing. *He* was not perceiving the moth *as* bark, moss, etc. because his perceiving had not even sorted out the moth as a *separate entity*. *I* can speak of his making a false perceptual judgment about a region that *I* can define within his non-epistemic seeing, but it is not, from *his* point of view that he has a false belief *about the moth*—he has no belief about the moth at all. He has not even sorted out that region that constitutes the moth for me; even the outline he has painted in his picture was not perceived by him, no more than are those cats in a child's puzzle picture which are 'hidden' among the lines making up branches and foliage perceived by the unwary child, as D. M. MacKay³³ has pointed out."

Spike Milligan can hear the irony in "pronounced limp" *before* the second context-indicator arrives in his utterance, but there is no irony at all for his hearer until that second context-indicator does arrive. The hearer *has* heard the irony-word *taken non-epistemically* before the second context-indicator arrives, but the hearer has *not* heard the irony-word *taken epistemically* before the second context-indicator arrives. Were Close to object that this is like his Aunt-Mannequin example

where the two gestalts have a common boundary (he might say that the hearer has a false belief about "pronounced limp"), it can easily be generalized to a pun that *lies across words*—say, the name of that British comedian of the Thirties, Nomo King, which, it is said, he picked up from the notice "NO SMOKING" painted across a double door that split the words between 'O' and 'K'.

The Irony Model for this second stage of *perception*, can be applied to this camouflage case as follows:

1. the *non-epistemic seeing of a common region of matter* by the two agents (Wright and the entomologist)—and recall that this non-epistemic seeing is *itself* open to a first-stage Irony Model analysis, as already shown;

2(a). *context-indicators A* (for the entomologist): kind of tree, height from ground, type of discoloration, etc.—leading him to perceive

2(b). *percept A*: a moth of a particular genus;

3(a). *context-indicators B* (for Wright): tree, familiar countryside, time of year, etc.—leading him to perceive

3(b). *percepts B*. moss, lichen, fungi, dry and damp bark in random profusion.

Does an agent have to make a perceptual judgment of some sort or other? In answering this we must be careful how we define *agent*. Suppose a person blind from birth who *cannot* have his sight restored by removal of cataract is given his first visual experiences, flashes, say, by means of electrical stimulation of his cortex. This is plainly a case of non-epistemic seeing. No one would claim that he was seeing the electrical stimulus, or the current flowing through the wires to his head, or even the switch closed by the doctor which produced the stimulus—this pointless chain could only be interrupted and given perceptual meaning by the doctor and the patient agreeing on something. If the doctor says that the switch will be closed every time someone goes through the door, and he then places the switch under the mat by the door, we might now say that the blind man's non-epistemic flashes had become epistemic. The blind man might even say, "I *saw* someone at the door." His flashes are thus both epistemic and non-epistemic. They are visual experiences (non-epistemic) but they are now related to human purpose as it tries to cope with the contingencies of the flux (epistemic). These considerations lead

³³ D. M. MacKay, "Language, Meaning and God," *Philosophy*, vol. 47 (1972), pp. 1-17.

us to say that it is not logically necessary for us to make perceptual judgments of some sort, although perhaps most of us do most of the time.

This is not to deny, by the way, that human agents and some animals probably have some specialized selector-mechanisms already built into their sensory systems, making them already geared to respond to certain patterns in their non-epistemic sensing rather than others.³⁴ This does not affect the general nature of the philosophical argument. We might even say that at some points in the evolutionary scale, some creatures were born with pure non-epistemic seeing and some with instinctually organized sensing, that is, with some favoured pattern-responses built in. Those that did not have the response built in certainly died at a faster rate than the others, and so did not succeed as species. But this does not affect the argument. In any case, we have shown that non-epistemic seeing is still present in "raw-feel" experience; it remains logically prior, and that is all we need for the argument here. As Kant wrote: "If we admit an object, [we] must admit it under the conditions of sensible intuition" . . . "Thoughts without content are empty, intuitions without concepts are blind"³⁵—but "blind" is only a metaphor for lack of *perception*, not lack of *intuition*.

Notice that we used the term "perceptual judgments" rather than "perceptual beliefs" at the end of the paragraph before the last. If someone started to fool the blind man by interfering with the circuit, he might not be able to attach belief to his perceptual judgment that someone was at the door at each flash. What is more, it is logically possible for him to have mental imagery of such flashes, to have an hallucination of those flashes. How does he know what to believe? We are back with our Cartesian question, and, having applied the Irony Model twice over, we now apply it a third time to effect an escape from this dilemma.

If it is logically the case, as the Three-Screens example has shown, that there is no inherent distinction between the sensory experiences of mental imaging, 3-D picturing, and open-eye-viewing, how can we escape from the sceptical trap? Just to remind you of the disturbing implications of the Three-Screens example in a more disturbing way

still: let there be *one* screen and let our logically possible experimenter gradually replace on the same screen the 3-D picture, the transparency effect (revealing the real billiard ball behind), and the externally controlled mental image or hallucination. If he did it so gradually and so carefully that no changes in luminosity or positioning took place, he could change from one to the others just as he liked—even mingle all three in varying proportions—and the subject would report no alteration at all! He is a cleverer demon than the one conjured up by Descartes. The alteration from 3-D picture to exactly matching reality behind the screen is well within empirical possibility now; the subject would not be able to tell when it was purely one, or half the intensity of one and half that of the other, or a quarter of one and three-quarters of the other—the gradations are infinite between "real" and "picture," and yet the subject would not be able to detect any change. R. N. Haber³⁶ has already performed experiments producing interaction between pictures and *eidetic images*, so perhaps all three could be managed after all by some ingenious psychologist.

So the solipsistic reef must be navigated. It is no use trying to drop anchor—or even sail back, pretending that non-epistemic sensing does not exist, for we have firmly established that it does. Moore was right.³⁷

Let us remember R. J. Hirst's attack on the "TV Theory,"³⁸ that there was no way in which we could "get into the studio to see the actors and scene in front of someone else's camera." Adopting a representational theory, he says, lands us straightaway with the objectionable "permanent possibilities of sensation," or worse, "things-in-themselves." Let us see whether they are as objectionable as they are made out to be, and whether it is even right to talk of "they"—even though something may exist where we think that "they" are.

Suppose now that two agents are *agreed* on a percept. This is quite unlike the camouflage case, where there was only agreement on the non-epistemic seeing (plus *some* epistemic, such as the ground, the sky, the fields behind, *most* of the tree—although I had thought that it was all of the tree). For the camouflage case we can already argue that

³⁴ Peter K. Machamer, "Recent Work in Perception," *American Philosophical Quarterly*, vol. 7 (1970), pp. 1–52 (see p. 8).

³⁵ Immanuel Kant, *Critique of Pure Reason*, B335, B75.

³⁶ R. N. Haber, "Eidetic Images," *Scientific American*, vol. 220 (1969), pp. 36–44.

³⁷ G. E. Moore, "A Reply to my Critics" in *The Philosophy of G. E. Moore*, ed. by Paul Schilpp (La Salle, Illinois, 1968), pp. 535–677.

³⁸ G. M. Wyburn, R. W. Rickford, and R. J. Hirst, *Human Senses and Perception* (Ontario, 1964).

the entomologist has been virtually able to "go into my studio," for he has been able to point out to me something already there which I had not myself perceived. But what of the normal agreement case? Let us say that this percept is some beads on a table. Both agents are prepared to call them "beads". They move to a second table: this time Agent 1 says, "Here are some more beads," but Agent 2 contradicts him—"There are no beads here!"

This little scene could have a great variety of explanations, but we shall content ourselves with two only:

- I. Table A: a necklace of brightly coloured wooden beads;
Table B: a necklace of irregular lumps of lignum vitae.
- II. Table A: a rosary;
Table B: a necklace of brightly coloured wooden beads.

I. The conflict in this case has come about because there is a semantic difference between the two agents, their modes of bonding their percept to their non-epistemic seeing are different. Whenever Agent 2 had been taught the word "beads," he had always had the sensory experience of something brightly coloured, shining, sparkling. He had never consciously analysed his bond of percept-and-sensing; it had been an immediate response with him. But now an irony reveals itself. On former occasions the ambiguity between himself and others had never come into the open. Behavior had converged to sustain apparently common intentions. Now he is faced with something dull, matt, opaque. He refuses to apply the learned percept, although Agent 1 does, having been more impressed in the past in *his* bond of percept-and-sensing by what we would call—though he had not, having never consciously analysed his bond—a pattern of roundish shapes linked by a thin line, together with other responses. How is this conflict resolved? By appeal to *wider intentions*. For example, do the agents live in a culture that frowns on aesthetic novelty or one that frowns on boring traditionalism? Whose *interpretation* of the percept wears the trousers? Suppose it was a culture that frowned on boring traditionalism; suppose Agent 1 to be the older of the two, with greater social prestige, and having considerable powers of rhetorical persuasion. He might then and there persuade Agent 2 to alter his bonding of percept-and-sensing. Hence, Agent 2 is now able to re-

inspect his mental images (he is a good visile) and perceive beads where he did not before. He is now able to see objects as beads that he never recognized before. His world has been changed by an Other, in its past as well as its present. What is more, he finds that this little persuasion, in company with many others of a similar kind from Agent 1, leads him to greater satisfaction of his intentions, both of physiological and cultural needs (and these, of course, are not exclusive). There might be sense now in saying that Agent 2 has been shown what was in Agent 1's "studio" at least as regards epistemic seeing; and, since it is logically possible for him to *re-inspect* his own mental images and get a new retrospective perceiving of them, we might say that he has learned something indirectly of Agent 1's non-epistemic seeing, for there was something in front of Agent 1's "camera" that Agent 2 is taught to notice that he had not examined before *even though it was in front of his "camera."*

II. The conflict in this case has probably occurred. Imagine a scene: it is in a monastery of the thirteenth century; our two agents are an abbot and a novice monk. At Table A they both agree to call the rosary "beads": at Table B, the novice monk (Agent 1), says that the necklace is an example of beads, the abbot refuses to countenance it. It is a fact of word-history that *beads* moved metonymically from meaning "prayers" to meaning "rosary," and from thence metaphorically to the necklace sense. We could ask again "Whose interpretation of the percept wears the trousers?—or rather, in this case, the *habit* of the cleric in charge?" If the abbot were a strict and puritanical man, he might be horrified at this near-blasphemous shift of meaning: if the novice were a budding artist with a sensuous eye for colour and pattern, he might object to the limitation being placed on his utterer's meaning. At Table B he would *perceive* beads as a beautiful string of patterned spheres: the abbot would not, and yet both would have a similar non-epistemic viewing. At Table A the Irony Model applied, *although neither of them was aware that it did*. Here is the analysis for their viewing at Table A:

1. *sensory presentation* (to which the non-epistemic five-fold analysis already applies, for each of them has a separate non-epistemic subjective viewing in his own body);

2(a). the *context-indicators A* (for the abbot): the crucifix on the rosary, his past uses of the rosary for prayer, the numerical pattern of the

ten "Hail Marys" and one "Our Father" repeated in the beads, and so on—leading him to perceive

2(b). *percept A*: a rosary.

3(a). the *context-indicators B* (for the novice): the beautiful patterns and colours that he readily links with those of craft-work he has done, his memory of artistic satisfaction in such craft-work, the delight others have shown on being given the *beads* he had made, etc.—leading him to perceive

3(b). *percept B*: beautiful small spheres strung on string for decoration of dress.

What could be the outcome of this conflict at Table B? It is not as simple a matter as Wittgenstein's "Continuing a Series"³⁹ or St. Augustine's *ambulare*.⁴⁰ Surely it would depend again upon whose persuasions reached more deeply, more widely. Of course, the abbot might appeal to force, which is a way of directly interfering in the immediate surface *intention* of the Other. Whether it had a more permanent effect would depend on the intention debates within the wider moral community. Perhaps it might be the case that the punished novice comes round to accepting the abbot's puritanism, in which case he will happily admit that he was "forced to be free." Notice how the percept is related to the intentional perspective of the individual agents. They proceed in what they regard as mutual knowledge, but which cannot be other than a presumed behavioural convergence. The Irony Model obviously applies to intentions as revealed in any perceptual rule used in common with another, and the ambiguity can interfere with the presumed common purpose at any time when the intentions, *taken for granted* as being in harmony up to that unexpected clash, turn out not to be so. This application of the Model well bears out the points made by Fred I. Dretske⁴¹ in his chapter on perceptual relativity, and the work done on utterances by Ragnar Rommetveit.⁴²

To teach somebody a percept, or to persuade him to shift it to a new application is to alter his intentional perspective, his set for future action. Therefore, to alter a percept is to alter in some

degree a *self*. This can only be accepted as satisfying to him if, in the long or short run depending on the circumstances, that self finds his intentions, both those openly known by him and those *taken for granted* by him, are given satisfaction. Of course, if his refusal to be so persuaded does in time produce failure of satisfaction and even increase of suffering, then he may at last swing his percepts towards those of his persuader, but the principle is the same. The only confirmation we have of the perceptual reality of the external world comes from the gestalt-shocks administered by Others to the Self, and, equally, by the Self to Others, shocks that prove themselves to be knowledge by their leading to the satisfaction of intentions as evidenced in such mutually guided behaviour; those intentions must include those taken for granted, and this implies a special difficulty since we can never be sure what intentions our opposite number is taking for granted. Yet this is the ground of objectivity and truth, this discovery of an ambiguity in what we had before epistemically perceived in confidence: we are thus able to acknowledge an independent external source—that we are non-epistemically sensing—of the Self's and the Other's experiences. It is not so much our percepts that are proved to be objective, as *what our percepts sort*. It is not so much a world of *objects*—and that is why it is confusing to talk of well-known material objects "causing" perception—but it is a world of *matter* upon which we do have, therefore, for most of the time, a direct view of our own. We do sense matter non-epistemically, but it is only because of the discovery of Others' differing perspectives which we retrospectively affirm to be correct OR our own perspectives which Others retrospectively affirm to be correct that we can claim this to be so. As I have said elsewhere,⁴³ Plato was wrong—in his Cave, it is not the shadows on the wall we see, it is the *wall*, shadowed.

The first philosopher that I know of to point a way to this refutation of solipism was Ludwig Feuerbach:⁴⁴

An object, a true object, is given me only when a being that affects me is present, when my own activity—if I

³⁹ Ludwig Wittgenstein, *Philosophical Investigations*, tr. by G. E. M. Anscombe (Oxford, 1967), pp. 57ff.

⁴⁰ St. Augustine, *De Magistro*, tr. by Robert P. Russell in *The Teacher; The Free Choice of the Will; Grace and Free Will* (Washington, 1968), pp. 14–15.

⁴¹ Dretske, *op. cit.*

⁴² Ragnar Rommetveit, *On Message Structure* (London, 1974).

⁴³ See the author's "Illusion and Truth," *op. cit.*

⁴⁴ Ludwig Feuerbach, *Die Philosophie zur Zukunft*, ed. by H. Ehrenberg in *Frommanns Philosophische Taschenbücher* (Stuttgart, 1922), p. 68.

proceed from the standpoint of the thinker—finds its limit, its resistant, in the activity of another being. The idea of an object is originally nothing other than the idea of another I.

and one remark especially apt:

what I see myself I doubt: only that is certain which another sees.⁴⁵

He anticipated Sartre's attempt to steer over the reef of solipsism, that of the re-making of the Self through *shame*:

My fellow-man is my objective conscience; he makes my failings a reproach to me; even when he does not expressly mention them, he is my personified feeling of shame.⁴⁶

Sartre's claim is that we find ourselves in the Other's world as a result of his "look," which brings shame:⁴⁷

The Other's look makes me be beyond my being in the world and puts me in the midst of the world which is at once *this world* and beyond this world.

This acknowledges the *irony* of the case when, through becoming aware of the Other's gestalt upon our own Self and therefore our interpretation of the world, we shift to a new intentional perspective on our own Self in all its non-epistemic experience. We discover an ambiguity in ourselves, in what before we perceived and conceived with the most personal of confidence. The Other has provided deeper *context-indicators* of our own intentions that before we had not recognized. We became conscious of an irony of which before we had been completely unaware: we recognize the Other's intentional perspective, see it deeper and more far-seeing than our own in revealing to consciousness what had been taken for granted; the ironic pattern grows clear, and we shift our perspective on the same experience to a new one. And yet—and this is where Sartre is in error—we might be mistaken in this! Shame is *not* an infallible guide: we can even learn to be ashamed of a shame. Still, we might feel that we can say with some security: if an experience with an Other in a supposed dream can alter the nature of the dreamer and all his past dreaming, it cannot be a dream! This is the strong core of Feuerbach's and Sartre's arguments. It is strange though that Sartre did not

add what is equally true, that I become convinced of the objectivity of a discovery *of my own* when I myself effect a gestalt-shock of Self- and world-re-ordering upon an Other, when I induce a Self-transformation in him *that he willingly concedes*.

Feuerbach's and Sartre's way with solipsism, however, does not answer all objections. If non-epistemic sensing is common to imaging, picture and real, and no distinction can be made between them on the grounds of vivacity or steadiness or control by will or immediate coherence—on all of which Berkeley depended—why should the satisfying of intentions have such epistemological value? The hardened sceptic might ask us this awkward question: "Why shouldn't you dream intentions too? Why shouldn't you dream hunger and thirst and the eating and drinking that satisfy them? If this is your only criterion, it is just as open to the solipsistic accusation! People can have hallucinations of sight, hearing, the kinesthetic sense, yes—but they can also have hallucinations of pain, hunger, thirst. Aren't you and those you may have convinced now in desperate case with no proof of the external world at all? Your 'gestalt-shocks' to percepts and even to the Self-concept depend on the ultimate satisfaction of intentions as you have yourself admitted. Why shouldn't the very experiencing of them be illusory?"

First, intention is not only a matter of gaining gratifications and avoiding suffering for oneself and for others, it is also a maintaining of life and avoidance of death. It seems to me reasonable to suppose that, if my non-epistemic sensory experience was an incalculable mixture of dream and real so that I could not tell when I was feeling hungry and really eating, then I would have probably died by now. There are examples of "psychological solipsists," schizophrenics who have become deprived of self-hood;⁴⁸ some of them have died as a result of a failure to distinguish dream and real. The fact that I am still alive appears to be one good inductive ground for belief in there being a material base for my perceptual judgments, however much they remain open to revision. The threat of death is very real to me: I often dream about it.

Secondly, at the risk of begging the question, I feel I should ask our hardened sceptic to help me

⁴⁵ Ludwig Feuerbach, quoted in Eugene Kamenka, *The Philosophy of Ludwig Feuerbach* (London, 1970), p. 101.

⁴⁶ Ludwig Feuerbach, *The Essence of Christianity*, tr. by George Eliot (New York, 1957), p. 158.

⁴⁷ Jean-Paul Sartre, *Being and Nothingness*, tr. by Hazel E. Barnes (London, 1966), p. 261.

⁴⁸ W. Donald Oliver, "A Sober Look at Solipsism," *American Philosophical Quarterly Monograph Series*, No. 4 (1970), pp. 30–39.

understand his *speaker's* meaning when he says, "All your world is a dream." If he means that dream and real share the character of being non-epistemic, or that all our percepts are products of our imagination working in the world, there I agree with him, but if he is really trying to tar me with the brush of "metaphysical solipsist," then what does he mean by it if he accepts the truth of his own accusation? Does he suggest that he is *himself* a figment of this extended illusion I am having? In that case, to accept his own argument is paradoxical, for it is at once to discount him and it as illusory! And then, since meanings are established across agents by behaviourally recognized contrasts and comparisons, as Bambrough has reiterated, what contrast that he shares with me can give meaning to his word *dream*? If he rejects reality, what does *dream* mean when he utters it because it is obvious that he has nothing to contrast it with. He is illicitly trading on a distinction between dream and reality to ask his question. It is he who is begging the question, not I.

Thirdly, since it is a being who is sensing in a non-epistemic way he is questioning, why doesn't he go and ask his question of a uni-cellular animal with a single light-receptor? His question would make as much sense from this point of view: the creature is certainly sensing non-epistemically, so shouldn't he be wondering whether the whole of reality, including himself, is compassed in the subjective experience of such a creature, and that there is no material world at all, and all that exists is nothing but that creature's sensory experiencing? It seems as much sense to wonder whether such a uni-cellular animal with its single receptor is a metaphysical solipsist as to wonder whether I am or any of us is.

There is one final objection from another quarter which must be dealt with before concluding. The realist is likely to ask us, "I understand us having hallucinations of blue or of the smell of smoke, but I cannot understand an hallucination of *pain*. Surely this is the best defence against the hardened sceptic. Our experience of pains and pleasures is our surest foundation for objectivity and truth because it is purely objective. A pain is always a pain: it is never an hallucination of one. Why not claim epistemological value for pains and pleasures, and hence *intentions*, on this account?"

This is a defence against the sceptic that I am not prepared to take up precisely because it denies the non-epistemic in the case of pains and pleasures. A book in front of me now I am prepared to call

"blue"; I am fixing my specially learned percept to the non-epistemic seeing, which actually consists of a quite unnameable *continuum* of shades. That unnameable continuum, which I can also mental-image, is part of my non-epistemic visual field (a part which I am artificially separating off merely for the purposes of argument). Whether what I am non-epistemically seeing is real or mentally imaged I have to establish *on other grounds* than those of its sensory character; there is no doubt that I am having a *vivid* non-epistemic experience in *both* cases, real and imaged. Now the case is no different with pain. If the amputee experiences a pain in his toes, his non-epistemic experience is exactly the same as when he had a leg, but he has to establish *on other grounds* whether he has real experience or not. In fact only the doctor may know at that moment that the toes in question are not there. The amputee has to ask the doctor whether his pain is real or not for he has a *vivid* non-epistemic experience in either case. The question makes perfectly good sense, and is no different from the question "Am I seeing a real blue or not?" for the person who, like Macbeth, is wondering *whether he is having an hallucination* because he is having a vivid non-epistemic experience just the same.

So this last argument brings us round to the "I-Thou" once more, to Feuerbach's "what I see myself I doubt: only that is certain which another sees," but we must conclude on a vital correction of this very maxim. Feuerbach was wrong to think that society, the Other, guaranteed truth and objectivity. Dialectic is a *two-way, ironic* process. Sometimes the Other corrects Me, and most often because He—as all ancestors and mentors—has shaped Me: but sometimes I correct Him, and all my ancestors and mentors. Feuerbach's maxim can now be seen to be in part a *pun*. "Only that is certain which another sees" could be taken in a way Feuerbach did not expect: it is not only that I rely for truth and objectivity upon what the Other shows Me *with the gestalt-shock that He administers to Me*, but that a discovery of truth that I make becomes certain when I convince the Other of it, *when I administer a gestalt-shock to Him*.

The forming of new percepts and the application of old ones can always be called a *judging*, but it cannot be called a *believing* until I and the Other have tested out our own uses together to see whether intentions are satisfied. This is because, whether openly known or taken for granted, consistent with each other or inconsistent, the intentions of a socialized self cannot but involve the

III. FOUNDATIONAL VERSUS NONFOUNDATIONAL THEORIES OF EMPIRICAL JUSTIFICATION

JAMES W. CORNMAN

IS an empirical statement known only if some particular set of statements forms the evidential foundation which provides the ultimate justification of the statement? For many philosophers over many years, the answer to this question has been almost dogmatically affirmative. Recently, however, strong and diverse attacks have been launched against a foundational theory of justification by a wide variety of philosophers. At present, R. M. Chisholm stands out, almost alone, as the embattled defender of foundationalism. One of my primary aims in this paper is to examine that part of Chisholm's defense that consists in an attack on all nonfoundational theories. Before this examination, however, I must identify the basic foundational thesis and its major non-foundational opponents. I shall begin by characterizing the minimal version of each sort of thesis. This is important because a thesis is refuted only if its minimal version is refuted. Thus we need the two minimal theses in order to evaluate both Chisholm's objections to alternatives to foundationalism, and attacks of others on foundationalism. It will also, however, help us classify more precisely the theories of philosophers, such as Quine and Sellars, who are all too quickly described as non-foundationalists.

1. Chisholm's Thesis and Minimal Foundationalism

Because we shall consider Chisholm as a paradigm of a foundationalist, let us begin our search for a characterization of minimal foundationalism with a view he endorses that he calls "thesis A":

A Every [empirical] statement, which we are justified in thinking that we know, is justified in part by some [empirical] statement which justifies itself.¹

To see whether thesis A adequately characterizes

minimal foundationalism we should unpack two phrases: " x justifies y in part," and " x justifies itself." In both cases I shall assume the phrases are to be relativized to a person and time, and in the first case I shall assume that x must be a needed part of the justification of y . Thus:

D1 x justifies y in part (at least) for person s at time $t =_{df}$ there is evidence, e , that is a set of statements which justifies y for s at t , and e would not justify y if it did not contain x either as an element or as a conjunct of a statement that is logically equivalent to an element.

In the second case, it might seem we need only require that x is justified even if nothing else is, but this will not do because self-justifying statement, x , might justify another by entailing it. Then it would not be possible that x alone is justified. Let us use, then, the following:

D2 x is self-justifying (justifies itself) for s at $t =_{df}$ x is justified for s at t , and it is false that it would be justified for s at t , only if it were to be justified by some relationship it has to some other statement or group of statements.²

Note that *D2* is consistent with the view that a sentence is self-justifying for a person only at times when certain conditions are met. That is, *D2* does not require that a sentence be self-justifying under all conditions and at all times. The crucial point is whether or not some other statement is needed to justify it for a person at that time. For example, it might be that the report, "I am now having a visual experience of my holding something large and heavy," is self-justifying for s at t only if it does not conflict with any other report that s believes at t , such as "I am not now having a tactual experience of something large and heavy."³

¹ Roderick Chisholm, "Theory of Knowledge" in R. Chisholm *et al.*, *Philosophy* (Englewood Cliffs, N. J., 1964), p. 263.

² *Ibid.*, p. 273, for Chisholm's different definition.

³ For a more detailed discussion of related points, see my "On the Certainty of Reports about What is Given," *Nous* (forthcoming).

Let us now use $D1$ and $D2$ to replace A by the following:

$F1$ Any empirical statement, p , is justified for s at time t , *if and only if* either p is self-justifying for s at t ; or (i) some evidence, e , justifies p for s at t , and (ii) e contains some empirical statement, x , other than p , that is self-justifying for s at t , and e without x does not justify p for s at t .

Note that $D2$ and $F1$ require that x be a statement, and they allow that something other than statements—for example, a certain experience—is required to justify a self-justifying statement. Should a statement justified only by an experience be called self-justifying? We could change $D2$ so statement x would be self-justifying only if nothing of any sort should be needed to justify it. I propose, however, that we should leave $D2$ unchanged for the minimal foundational thesis. I believe all foundational views require there to be self-justifying statements in someone's evidence, but the view that a statement is self-justifying only if *nothing* at all is needed to justify it, is too strong to express a necessary condition for the minimal view. What is required by clause (ii) in $F1$, when unpacked by $D2$, is strong enough.

I find that $F1$ states neither a necessary nor a sufficient condition for any foundational theory, and so it is not the minimal thesis. There are three reasons why it should be amended. First, on the assumption that $F1$ is a material bi-conditional, it would follow from universal skepticism about justification. But skepticism should not be allowed to justify the thesis. Second, $F1$ allows some statements to be justified for someone without any recourse to self-justifying statements. According to $F1$, a person need only have available *some* set of evidence for p that requires him to rely on a self-justifying statement. But he need not use it to justify p . Such a situation is inconsistent with every foundational theory, because each one requires a person to have recourse to some self-justifying statement at some stage of his justification. Third, $F1$ requires that some set of statements, e , which justifies p for s , contains a self-justifying statement. But consider a series of sets of statements, e_1 through e_n , where each e_i in the series justifies e_{i-1} , but only e_1 justifies p . Any foundational theory should require that *some* e_i in any such chain of justification for p contains a self-justifying statement, but it need not

require that, in at least one of these series, a self-justifying statement is a member of e_1 , the set of statements that justifies p .

To amend $F1$ appropriately, let us use the concept of "evidential series" characterized as follows:

$D3$ E is an evidential series (E-series) for $p =_{df}$ E is a series of evidence sets, $e_1, e_2, e_3, \dots$ such that: (1) e_1 justifies p ; and (2) for any e_i in series E , e_i justifies e_{i-1} in E .

In other words, any such a chain of justifiers, from e_1 to some first justifier or *ad infinitum* is an E-series for p . And we can use this notion of an E-series to define another term that will be useful later:

$D4$ x is an evidential ancestor (E-ancestor) of $p =_{df}$ there is an E-series, E , for p , such that for some set e_i in E , e_i would not justify e_{i-1} if it did not contain x either as an element or as a conjunct of a statement that is logically equivalent to an element of e_i .

That is, roughly, x is an essential part of some evidence in a chain of justification that ends with evidence justifying p .

We can now amend $F1$ to give us the following replacement:

$F2$ Any empirical statement, p , would be justified for s at time t *if and only if* it would be true that: either p is self-justifying for s at t ; or (i) there is an E-series for p that justifies p for s at t , and (ii) all of s 's E-series that justify p for s at t contain some empirical statement that is self-justifying for s at t .

I believe that $F2$ is minimal, that is, all and only theories that satisfy it are foundational. Of course $F2$ is not sufficient for many species of foundational theory. Chisholm, for example, seems to hold a Cartesian-like foundational theory which states that, of empirical statements, only those exclusively about either what is phenomenally given or about certain psychological attitudes are self-justifying.⁴ Such a theory implies, but is not implied by $F2$.

2. The Minimal Nonfoundational Theory

The basic opponent of foundational theories is often called the coherence theory of justification. But I think it better to view the coherence theory as just one of many nonfoundational theories, rather than its minimal version. I also believe that the minimal nonfoundational theory, should re-

⁴ *Ibid.*, pp. 273-285.

quire no more than that no self-justifying statements are *needed* for justification, so that it does not matter whether there are self-justifying statements. However, the minimal thesis should apply to every non-self-justifying statement. Thus we can rule out the contradictory of *F2*, which is true if merely *one* non-self-justifying empirical statement should be justified without reliance on self-justifying statements. The minimal doctrine should state that no justified, non-self-justifying empirical statements require justification by a self-justifying statement. Let us try then:

N Any empirical statement, *p*, that is not self-justifying for *s* at *t* would be justified for *s* at *t*, *if and only if*, it would be true that there is an E-series that: (1) justifies *p* for *s* at *t*; and (2) contains no empirical statement that is self-justifying for *s* at *t*.

Thesis *N* allows there to be self-justifying statements, and it allows them to justify other statements. It only requires that none of them are needed to justify non-self-justifying statements. It may be objected, however, that no nonfoundational theory should allow self-justifying statements to justify other statements. It might seem, then, that we should add to the right hand side of *N* that none of *s*'s E-series contain any empirical statement that is self-justifying for *s* and that justifies *p* in part for *s*. Nevertheless, I propose *N* as the minimal theory, because any theory stating that every empirical justification would be accomplished without self-justifying statements—so it does not matter whether there are self-justifying statements—would be a nonfoundational theory. If so, then the minimal thesis should be no stronger than *N*. And it surely should be no weaker than *N*. And as *N*, the minimal thesis conflicts with the minimal foundational thesis, *F2*, in the sense that either *N* or *F2* is false if some non-self-justifying empirical statement is justified.

3. *Objections to Nonfoundational Theories*

With it settled that *N* is the basic alternative to *F2*, we can turn to Chisholm's attempt to show that the main nonfoundational alternatives to his thesis *A* should be rejected. I shall reconstruct his procedure as an attack on the chief varieties of *N*. There are two species of *N*: first, each justified non-self-justifying statement is justified by an E-series containing only justified E-ancestors but no self-

justifying statement; and, second, each of these is justified by some series containing an E-ancestor that is not justified in any way, perhaps even unjustified in the sense of its denial being reasonable. But we can ignore E-series with such unjustified statements, because no series requiring an unreasonable statement justifies anything.

The first species allows two important varieties. If we followed Chisholm, one would not permit any statement in the relevant E-series for *p* to be partly justified by *p*, and the other would allow some of these statements to be partly justified by *p*. The first is supposed to lead to an infinite regress and the second to a circularity.⁵ As stated, however, the first does not guarantee an infinite regress. According to it, each justified non-self-justifying statement, *p*, requires a series, *E*, of justified non-self-justifying statements, excluding *p*. Thus each statement in any *e_i* in *E* must in turn be justified by a set, *e_{i+1}*, of justified non-self-justifying statements that does not contain *p*. But this does not generate an infinite regress. Assume that *p* is justified by *e₁* containing statements *q₁*, *q₂*, *q₃*, which are justified non-self-justifying statements that are different from *p*. Assume also that *q₁* justifies *q₂*, *q₂* justifies *q₃*, and *q₃* justifies *q₁*. Here is a case that satisfies the first thesis without an infinite regress. Of course, it is bothersome that *q₁*, *q₂* and *q₃* are each E-ancestors of themselves, but that is allowed by this statement of the first variety of the first species. In other words, this version allows a certain sort of circularity which might better be included in the second variety. Consequently, to rule out any sort of circularity and to guarantee an infinite regress, the first variety should be stated so it allows neither *p* nor any E-ancestor of *p* in the series to be one of its own E-ancestors. Then the second would allow both sorts of E-ancestors.

4. *Three Varieties of Nonfoundationalism*

We have three main varieties of *N*. First:

N1 Any empirical statement, *p*, that is not self-justifying for *s* at *t* would be justified for *s* at *t*, *if and only if* it would be true that there is an E-series that: (1) justifies *p* for *s* at *t*; (2) contains no empirical statement that is self-justifying for *s* at *t*; (3) contains only E-ancestors of *p* that are justified for *s* at *t*; and (4) contains no empirical statement that is itself (or is analytically equivalent to) one of its own E-ancestors.

⁵ Chisholm, *op. cit.*, p. 264.

Second:

N_2 Any empirical statement, p , that is not self-justifying for s at t would be justified for s at t , if and only if it would be true that there is an E-series that: (1) justifies p for s at t ; (2) contains no empirical statement that is self-justifying for s at t ; (3) contains only E-ancestors of p that are justified for s at t ; and (4) contains some empirical statement that is itself (or is analytically equivalent to) one of its own E-ancestors.

Third:

N_3 Any empirical statement, p , that is not self-justifying for s at t would be justified for s at t , if and only if it would be true that there is an E-series that: (1) justifies p for s at t ; (2) contains no empirical statement that is self-justifying for s at t ; and (3) contains some E-ancestor of p that is not justified for s at t .

We can reconstruct Chisholm's reasons for rejecting the main alternative to his A , as reasons to reject the three preceding theses. That is, he would reject N_1 because it leads to an unacceptable infinite regress of justification. He notes that this can be avoided by making "blind posits" or guesses. But to do this is to desert N_1 for N_3 , which he would reject because it fails to distinguish "between knowledge, on the one hand, and a lucky guess, on the other."⁶ Finally, it seems he would reject N_2 because it requires circular justification.⁷

5. *An Objection to the First Variety: Justification and Infinitely Long E-Series*

It is clear that N_1 requires an infinitely long series of E-ancestors for every justified non-self-justifying empirical statement. It also seems, initially, as Chisholm requires, that the regress is vicious, in the sense that whatever requires such a regress is unacceptable. Recently, however, B. Aune has argued that the required regress is not vicious, because only an unreasonable thesis about justification would make it vicious. As applied to N_1 it would be:

A series of E-ancestors of p justifies p for s only if s actually justifies, individually, every statement in the series that is not self-justifying for s .⁸

Contrast this with the thesis that requires only that s be able to justify, individually, any such statement in the series. The first thesis, conjoined with N_1 , leads to a vicious regress, because it assigns an impossible task to s . But the second thesis requires of s neither an impossible nor even an implausible task. He must merely be able to justify any particular statement in the series. He need not, then, either be able to justify or actually justify individually the whole set of statements in the series. Thus, according to Aune, the regress is not vicious, if, as he maintains, the first thesis about justification can be rejected for the second.

I agree with Aune in rejecting the first thesis. No more should be required of s than what the second thesis states. Furthermore, it seems plausible that if s actually justifies a significant sample of E-ancestors of p by other statements, and he does so in a way that gives reason to conclude he is able to justify non-circularly any one in the series by another, then it is reasonable to conclude that each statement in the series is justified by another, and that s is able to show this for any particular one of them. However, finding the infinite series of N_1 innocent of viciousness does not solve the main problem confronting N_1 , namely, that it seems to allow infinite E-series to justify sets of statements, where it is clear that some members of the sets should not be justified.

For any statement we can generate an infinite E-series that meets the requirements (2) through (4) of N_1 . And for any set, e_i , in the series we can easily formulate a set that justifies the statements in e_i . Consider the singular statement Pa , where a is not a number. This is justified for s at t by $e_1 = \{Q_1a, (\forall x)(Q_1x \supset Px)\}$. Then the statements in e_1 are justified by $e_2 = \{Q_2a, (\forall x)(Q_2x \supset Q_1x), (\forall x)[(Q_2a \cdot Q_1x) \supset Px]\}$. And those in e_2 are justified by $e_3 = \{Q_3a, (\forall x)(Q_3x \supset Q_2x), (\forall x)[(Q_3a \cdot Q_2x) \supset Q_1x], (\forall x)[(Q_3a \cdot Q_2a \cdot Q_1x) \supset Px]\}$. This series, E_1 , goes on to infinity, because we have an infinite number of predicates of the form Q_ix . For example, we can let $Q_ix = x$ is either not a number or is greater than the number i . This series contains no statement that is its own E-ancestor, and, we can assume, it contains no self-justifying statements. Thus E_1 meets requirements (2) through (4) of N_1 , and so, if meeting those requirements is sufficient for E_1 to justify p , then, by N_1 , Pa is

⁶ Chisholm, *op. cit.*, p. 268.

⁷ Chisholm, *op. cit.*, p. 264.

⁸ See Bruce Aune, "Remarks on an Argument by Chisholm," *Philosophical Studies*, vol. 23 (1972), p. 329.

justified for s at t . But should Pa be justified in this way?

To see the problem E_1 raises for N_1 , consider the requirement:

R A theory of justification is acceptable only if it does not yield that both a sentence and its denial are justified for the same person at the same time.

If a theory violates what *R* requires of acceptable theories, then it yields that some sentence and its denial are both justified. Thus each would be more reasonable than its denial. But this implies that each of these is also less reasonable than its denial, because each is the denial of the other. So a theory that violates *R* should be rejected.

The main problem for N_1 is that if it allows its conditions (2), (3) and (4) to be sufficient for (1), then N_1 violates *R*. Consider Pa and its denial, $\sim Pa$. We can construct an infinite E-series, E_2 , for $\sim Pa$ just by substituting $\sim Px$ for Px throughout E_1 . Thus, contrary to what *R* requires, both Pa and $\sim Pa$ would be more reasonable than their denials for s at t , if requirements (2) through (4) are sufficient for justification. Of course, the problem would be avoided if N_1 could be amended to state a fourth condition as needed for an E-series to justify p for s at t . But what might it be? Might it restrict s to the use of just one of each pair like E_1 and E_2 ? But it should not merely allow s to choose whichever one he wants. This would permit him to justify whatever he wants by substituting appropriately throughout E_1 . Obviously, an N_1 -theorist cannot rely either on uncovering the presence of self-justifying statements in justifying E-series, or on making blind posits of statements in the E-series he desires to justify some statements. And there seems to be no plausible coherence feature that would be found only in the justifying infinite E-series required by N_1 . However, N_1 might be amended, in accordance with *R*, so that an E-series does not justify a statement, p , if there is no reason to pick that series instead of its analogue that would justify the denial of p . But then, I suggest, we will find that few, if any, statements will have justifying E-series according to N_1 , because few E-series would meet all five requirements of N_1 as amended. Consequently, because at present there seems to be no plausible way to amend N_1 , we should reject the sort of justification it requires—at least until someone proposes a viable emendation.

6. *Objection to a Second Variety: Blind Posits and Justification*

With N_1 in doubt, two sorts of theses that avoid foundationalism remain. If Chisholm is right, that sort which allows some justifying statements not to be justified by anything at all, that is, N_3 , should be rejected because it involves "blind posits" and so fails to distinguish between knowledge and lucky guesses. It is not clear, however, that this criticism is accurate when directed at N_3 . It is true that if N_3 is correct, then each statement, p , that is justified for someone, s , rests in part on at least one statement that is a "mere guess," in the sense that it is no more reasonable for s than its denial. But this does not show that s has no evidence at all for p , as would be true if s were merely guessing about the truth of p . Thesis N_3 allows some evidence and reasoning for and against p . It merely requires that some E-ancestor of p be not justified. Of course, if circular justification and infinite E-series are prohibited, then, sooner or later, some E-series for p ends with a final set of statements, none of which are justified. But N_3 by itself does not require this.

The crucial question for N_3 is not whether a statement justified according to N_3 is a mere guess or blind posit, but whether any statement, q , that is not justified for s is an immediate E-ancestor of a statement, r , in an E-series that justifies a statement p , for s . When this question is put this way, it seems that the answer is negative and N_3 should be rejected. When we consider q merely as one of many remote E-ancestors of p , it is not clear that the lack of justification of q has any ill effects. But this impression changes once we realize that to be one of many remote E-ancestors of p is to be, probably, one of only a few immediate E-ancestors of r . And that raises the question of whether any statement can be justified by a set of statements containing and requiring a statement such as q .

Consider the following argument about such a q and r . For r , in a set e_i , to be justified for a person, s , by a set, e_{i+1} , of its immediate ancestors, its relationship to the members of e_{i+1} must be one of two kinds: either that of conclusion to premises, or that of a statement that is probable relative to evidence. If the relationship is the first sort, then r is a conclusion that is inferred, either deductively or inductively, from several premises, one of which, q , is required but is not justified for s . But then the set of premises do not justify r for s , and s cannot justify anything else using r in set e_i . Consequently, no series which contains such a q as a remote

ancestor of p is an E-series that justifies p , contrary to N_3 .

On the other hand, if such a q is considered to be contained in a set, e_{i+1} , of s 's evidence relative to which this r is probable [that is, $\Pr(r/e_{i+1}) > 0.5$], the question arises about what else, if anything, is needed for r to be justified for s by the set e_{i+1} . Some people require that the probability of the conjunction, c , of the members of the set e_{i+1} be 1, but I think that this is too restrictive. Yet, surely, at minimum this conjunction must be probable [that is, $\Pr(c) > 0.5$]. Nothing is justified for someone by evidence, even if its probability relative to that evidence is very high, if the evidence itself is not probable for him.⁹ Given this, and that c is a conjunction of s 's evidential statements, $f_1, f_2, f_3, \dots, f_m$, then, because $\Pr(c) = \Pr(f_1 \cdot f_2 \cdot \dots \cdot f_m)$, the following is true by the probability calculus:

$$\Pr(c) = \Pr(f_1) \times \Pr(f_2/f_1) \times \Pr(f_3/f_1 \cdot f_2) \times \dots \times \Pr(f_m/f_1 \cdot f_2 \cdot \dots \cdot f_{m-1}).$$

But, letting $q = f_1$, then for s , $\Pr(c) \leq 0.5$, because q is no more reasonable than its denial for s . Thus again e_{i+1} does not justify r for s . Consequently, for both relationships that a justified statement has to the set of its immediate E-ancestors, every member of that ancestral set is more reasonable for s than its denial, that is, it is justified for s . But if every immediate E-ancestor of each set of E-ancestors of a justified statement must be justified, then, contrary to N_3 , no E-series for any statement contains one of its E-ancestors that is not justified. Thus, N_3 , like N_1 , should be rejected.

7. *Objections to the Third Variety: Coherence and Justification*

I think that few nonfoundationalists would mourn the demise of N_1 and N_3 , because most of them probably subscribe to some sort of coherence theory of justification and this requires N_2 . Such theories are often associated with coherence theories of truth, and often rejected because they require a circularity of justification. But it would be a mistake to think that a coherence theory of justification has important implications for theories of truth, or conversely, and it is clearly debatable whether the required circularity is vicious, that is, whether any theory implying it should be rejected. Consider, for example, what I take to be the minimal coherence theory of justification:

N_{2a} Any empirical statement, p , that is not self-justifying for s at t would be justified for s at t , *if and only if* it would be true that there is an E-series such that: (a) clauses (1), (2), (3), and (4) of N_2 are true; and (b) for p and every statement in this E-series, either they are justified for s at t by being in a justifying set of statements, J , for s at t , or they are justified by the set J for s at t .

Like N_1 and N_3 , N_{2a} allows foundational sorts of justification, but it requires there to be a coherence justification for every justified, non-self-justifying statement. It is also compatible with a variety of theories of truth, including what is rather vaguely dubbed the "correspondence" theory. Our concern with N_{2a} , however, is to discover whether it should be rejected. If it should be rejected, then we have reason for dismissing N_2 , because, I find, N_{2a} is its most plausible variety.

8. *An Explanatory Coherence Theory of Justification*

One way to approach the problem of evaluating N_{2a} is to find a clear example of a coherence theory, and use it as a paradigm case. Although there have been few attempts to formulate such theories clearly, it is fortunate that K. Lehrer has recently provided a statement of such a theory, which he later rejects for his own subjectivist theory which is a variety of thesis N_3 . It is an explanatory coherence theory (*EC*-theory), which he and many others attribute to Sellars and Quine. For our purposes we can use a version that is simpler than the one Lehrer considers. It uses the following principle to determine whether a set is a justifying set for N_{2a} :

EC The set J is a justifying set for s at t , *if and only if* the members of J are:

- (1) the nonredundant explaining members of some system of statements, C , that has maximum explanatory coherence among those systems understood by s at t (that is, C is a "maximal" system for s at t), and
- (2) the statements explained by C .¹⁰

Crucial to *EC* is the notion of maximum explanatory coherence which we can define in much the way Lehrer does:¹¹

⁹ I discuss this in "On Acceptability without Certainty," *The Journal of Philosophy*, vol. 74 (1977), pp. 36-44.

¹⁰ See Keith Lehrer, *Knowledge* (Oxford 1974), p. 165.

¹¹ *Ibid.*

D5 System *C* has maximum explanatory coherence among the set of systems, *B* =_{df} (1) *C* is consistent; and (2) for any system, *x*, in *B*, if *x* is consistent but incompatible with *C*, then (a) *C* explains at least as much as *x*, and (b) it is false that *x* explains at least as much as *C* and also explains more things better than *C* does.

Even with definition *D5*, much remains unclear about maximum explanatory coherence. One problem is to determine conditions for when one system explains better than another, but let us ignore that problem except for noting two points. First, as used here, "*x* explains *y*" does not imply that either *x* or *y* is true, and truth is not a test for deciding which of two systems explains better. Second, "systematic" tests are relevant to how well a system explains. Following Quine, we can list some of these tests as refutability, simplicity, familiarity of principle, scope ("implies a wider array of testable consequences"), and fecundity ("successful further extensions of theory are expedited").¹²

A different unclarity in *D5* is, however, too important to ignore. It concerns what it is for *x* to explain at least as much as *y*. I find three initially plausible construals:

- (a) *x* explains at least as many statements as *y* does;
- (b) *x* explains at least the statements *y* explains;
- (c) *x* explains at least as many statements that are to be explained as *y* does, and at most as many that are not to be explained as *y* does.

Interpretation (a) should be rejected. On some plausible theories of explanation, if a system explains *p*, then it explains infinitely many other statements, because it also explains *p* ∨ *q*₁, and *p* ∨ *q*₁ ∨ *q*₂, and so on. Thus on such a view, any system that explains at least one statement explains as many as any other system.

Interpretation (b) seems to be the view that many scientifically oriented coherence theorists would accept, but it makes *EC* too strong even if the explanatory systems for someone are, implausibly, limited to scientific theories. At present there is no one scientific theory that explains at least what every other one explains, not even if we consider only those theories currently in vogue. If it be replied that we should conjoin all these

theories to get the maximal system, one problem is that the result may well be inconsistent, at least in its theoretical implications. But more importantly, there are many individually consistent scientific theories, such as Newtonian mechanics, which, although not in vogue, are not eliminated by *D5* using (b), and clearly no consistent system explains what Newton's theory explains and also what Einstein's theory explains. Thus, if nothing of what each theory explains is eliminated, then no statements are justified for anyone who understands both theories, because no system is maximal for him. This is surely too strong. The only reply is to say that one of the theories has been refuted, and so no maximal system is required to explain all that each explains. But then the problem for an *EC*-theorist is to show that one of the theories explains something that is not to be explained. It is clear, then, that (b) must be amended so it considers only those statements explained by *y* that are to be explained.

As stated, (c) does not require such an amendment, but it does require that there be some way to determine which statements are to be explained. A person should not be allowed to decide this arbitrarily, because, as Lehrer notes, he can limit the statements that require explanation to exactly those his favorite system explains very well.¹³ Thus for (c), (b), and also, I find, for any plausible construal of *D5*, the crucial problem for an *EC*-theorist is to find a nonarbitrary way to restrict the statements that are to be explained. To see how difficult this task is, consider one plausible way he might begin. He might begin by limiting the statements to singular, categorical observation statements. He can do this either by identifying observation predicates and singular referring terms and, thereby, observation statements, or, as Quine does, by considering a sentence to be "observational insofar as its truth value, on any occasion, would be agreed to by just about any member of the speech community witnessing the occasion."¹⁴ The latter technique allows many statements that seem very theoretical to be observational. But the resulting set is still too large. He might next divide this set into maximally consistent subsets, but this leaves him with many conflicting sets, not all of which are to be explained. The natural way to eliminate further is not available to an *EC*-theorist. That is,

¹² W. V. O. Quine, *The Ways of Paradox and Other Essays* (New York, 1966), p. 234.

¹³ Lehrer, *op. cit.*, p. 170.

¹⁴ W. V. O. Quine, *The Roots of Reference* (La Salle, Ill., 1974), p. 39.

he cannot limit the set to those statements that are confirmed by observation, individually and independently of any explaining theory. Such statements would be self-justifying, contrary to what *N2* requires.

He might propose, instead, that the desired set include all and only true observation statements. But then the following objection arises. Let us assume that either Newton's theory or Einstein's theory (but not both) explains only true observation statements. If this present version of *EC*-theory is correct, then for all we can determine, the Newtonian theory and what it explains might be justified for us, because it explains only what is true. Most of us, following what scientists tell us, would believe—mistakenly and with no means for discovering our mistake—that Einstein's theory is justified. And anyone who doggedly has maintained the Newtonian theory would be justified in his belief even though he has no way to support his view except to say that what his theory explains is true. This is surely an objectionable feature of a theory of empirical justification, and it should be avoided. It seems, then, that an *EC*-theorist needs something more than the mere (lucky) truth of observation statements in order to limit the set of statements to be explained. He needs some non-arbitrary means that is more reliable than a mere guess or dogmatic assertion for choosing which of such conflicting theories is more reasonable.

I can think of just one other way an *EC*-theorist might proceed. He might try to base his elimination on agreement rather than truth. For example, he might propose the following Quine-like principle:

OE An empirical statement, e_1 , of language L is to be explained on occasion, O_i , if and only if it is a singular observation statement and almost all members of speech community M (e.g., scientists who speak L) would assent to the truth of e_1 on witnessing occasion O_i .

But is the right-hand side of *OE* more accessible than mere truth for an *EC*-theorist?

Consider a person, s , who uses *OE* to justify the acceptance of an observation statement, e_i , that is explained by Einstein's theory but denied by Newton's. To succeed, he must verify a general statement, such as:

Gr Many members of class M would assent to e_1 on witnessing occasion O_1 .

Since *Gr* is empirical but not a singular observa-

tion statement, it would be justified for s , according to this *EC*-theory, if and only if either it is a non-redundant explanatory part of a system that is maximal for s , or it is justified by set J . But how is an *EC*-theorist to justify that *Gr* has either status?

I have found only one procedure available for an *EC*-theorist who relies solely on agreement via *OE* to justify what is to be explained on each occasion. He can first assume *Gr* and some of its kin, then use the set of observation sentences which they and *OE* pick out to determine which explanatory systems are maximal, and, finally, discover whether *Gr* either is in a maximal system, or is justified by a justifying set, J , or is neither in nor justified by J . If *Gr* is in or justified by J , then it is justified and so is the claim that e_1 is to be explained. If *Gr* is neither in nor justified by J , then neither *Gr* nor the claim that e_1 is to be explained is justified. However, our *EC*-theorist can always supply *Gr* an explanatory role, and thus justification, by using *Gr*, on occasion O_1 , to help explain e_2 = Scientist, a , assents to e_1 . Of course, it must be justified that e_2 is to be explained on occasion O_1 , but this can be achieved by assuming *G2* which is like *Gr* except that *G2* contains " e_2 " instead of " e_1 ." And then *G2* can be justified by having it explain e_3 = Scientist, b , assents to e_2 . And this goes on *ad infinitum*.

The preceding procedure for justifying *Gr* requires infinitely many explaining sentences and sentences to be explained. But that is not its problem. Like the problem for *Nr*, its problem is that it is unjustifiably arbitrary. We need only substitute " $\sim e_1$ " for " e_1 " in *Gr* and e_2 , and, keeping all else the same, $\sim e_1$ is justified for s at t , instead of e_1 . But, as was noted previously, any procedure should be rejected, if it allows someone to justify whichever one of a statement and its denial he desires to justify. Consequently, like the problem he faces if he relies on mere truth, an *EC*-theorist who proposes principles like *OE* that rely solely on agreement has no nonarbitrary way to justify that some set from among the many conflicting maximally consistent sets of observation statements is to be explained. Thus he does not escape the preceding objection to his theory by substituting agreement for truth.

9 Conclusion about Explanatory Coherence Theories

I believe that the serious problem facing an *EC*-theorist who relies on truth or on agreement is not unique. That is, I believe that any other way to justify the elimination of certain statements will

have similar problems. It is clear, however, that I have not established this more general claim. Instead let me offer a challenge to any *EC*-theorist: refute the following dilemma or reject your theory. An *EC*-theorist must find a procedure for selecting a set of statements that are to be explained. He has but two alternatives: allow a person to restrict the set arbitrarily, or provide nonarbitrary but accessible particular restrictions. But the first alternative is objectionable, because it allows a person to pick just those statements that his favorite theory explains. And, as just indicated, it seems an *EC*-theorist has no way to realize the second alternative.¹⁵ Consequently, all *EC*-theories should be rejected, either for a less objectionable theory or for skepticism.

I conclude that we should reject all *EC*-theories until this challenge is met. Of course, someone may prove me wrong, but my confidence in my conclusion is strengthened by another objection from Lehrer.¹⁶ Even if being a nonredundant explanatory member of a system that is maximal for *s*, is sufficient for being justified for *s*, it is quite implausible that if *s* has no such explanatory system, then no empirical statements are justified for him. It surely seems that some of *s*'s observation statements and certain of his psychological reports are justified, whether or not they are explained for *s*.

Lehrer raises a third objection that should be noted.¹⁷ As stated above, *D5* allows several incompatible systems to be maximal, and so, Lehrer would claim, wherever this occurs, an *EC*-theory results in incompatible statements being justified for one person at one time. But that is wrong. Furthermore, if Quine is (or at least used to be) right in claiming that scientific theory is underdetermined, then there always are two conflicting, maximal systems, even when each explains exactly what is to be explained.¹⁸ Thus if Lehrer and Quine are both right, there is another objection sufficient to refute *EC*-theories. And, we might note, this objection is the analogue of the crucial objection that supposedly defeats the coherence theory of truth. I

think, however, that neither of these objections is decisive.

10. On Coherence Theories of Justification and Coherence Theories of Truth

The defect in *N2a*, as spelled out by *EC* and *D5*, that gives rise to the preceding objection can be easily corrected by adding a clause to *N2a* to the effect that a person may use at most one of his justifying sets at any one time. It may be objected that this gives a person too much leeway. But, since each set includes only one maximal system and what it explains, it seems reasonable to allow *s* to choose whichever one he wants, because no such choice would ever result in *s* rejecting one justificatory system for another that is less reasonable. But, although this reply helps *EC*-theories, it is of no use to coherence theories of truth. Not only is it most implausible to allow someone's desires to decide what is true, but also, because different people have different maximal systems and different desires, determining truth in this way would probably result in some statements and their denials being true, unless truth is relativized to persons. But although justification should be so relativized, truth surely should not.

11. On a Coherence Theory of Truth With a Foundational Theory of Justification

It might seem that we have found a place where coherence theories of justification are superior to such theories of truth. However, I believe a "truth-coherence" theorist can begin a solution of his problem in a way that allows him to avoid the two crucial objections to "justification-coherence" theories, namely, they have no way to justify any particular restriction on what is to be explained, and explanation is not a necessary condition of justification. Thus his thesis may turn out to be the more plausible of the two. Once it is realized that explanation is one, but only one, important means of justification, it is quite natural and very tempting to adopt a theory of justification that goes, very

¹⁵ Nicholas Rescher argues for a coherence theory of truth in *The Coherence Theory of Truth* (London 1973), and it might be thought that he provides a coherence theory of justification. What is relevant is his attempt to eliminate all but one of the maximally consistent subsets of what he calls "truth-candidates" of "data" (pp. 53-70), so that the remaining one is justified as the set of true propositions. However, in chapter V, where he suggests five ways to eliminate competing sets, the most plausible ways depend on independently establishing claims about probabilities, or plausibilities, or specially designated theses. But if what I argue in this paper is correct, some sort of foundational theory of justification is needed to establish such claims.

¹⁶ Lehrer, *op. cit.*, pp. 178-180.

¹⁷ Lehrer, *ibid.*, pp. 181-182.

¹⁸ See W. V. O. Quine, *Word and Object* (Cambridge, Mass. 1960), pp. 21-22. I discuss this in "Reference and Ontology: Inscrutable but not Relative," *The Monist*, vol. 60 (1976). For Quine's later views where he expresses doubts about underdetermination, see his "On Empirically Equivalent Systems of the World," *Erkenntnis*, vol. 9 (1975), pp. 313-328.

roughly, as follows. Singular, categorical observation reports (such as "I see something brown"), and, I would suggest, psychological reports (such as "I am hurting," and "I believe I see a table") can be tested individually and either confirmed or disconfirmed by observing or experiencing what occurs on the appropriate occasion. Those that are so confirmed are to be explained. Then the empirical sentences that are justified for a person *s* are those singular observation and psychological reports he has confirmed, plus any universal and statistical observational generalizations inductively confirmed by these singular statements, plus the sentences of one explanatory system that is maximal in sense (c) for *s* regarding these first two sets, plus what these three sets jointly justify. Of course, what results is not a coherence theory. It is instead foundational, but that by itself is no cause for alarm, not even for a "truth-coherence" theorist.

A truth-coherence theorist can adopt the preceding foundational theory of justification and combine it with a coherence theory of truth. In this way he avoids the two objections to *EC*-theories. And he might begin to avoid the preceding objection to his theory of truth by adopting something like the following rough thesis. The true empirical statements include those observation reports and psychological reports that would be individually confirmed at each moment by an all-perfect being, plus the nonredundant members of the explanatory system that would best explain all these reports for an all-perfect being, plus whatever is entailed by the first two sets of statements. Such a thesis would not be trouble-free, but it might be refined to avoid damaging objections, such as the previous one that no coherence theory of truth specifies exactly one consistent set of sentences as true. To avoid this crucial objection, this first, rough thesis about truth must be supplemented by some procedure for picking out exactly one member from each partition of empirical sentences that are not in the first three sets specified for an all-perfect being. Of course, if explanatory theory would be underdetermined for an all-perfect being, then the objection becomes even more difficult to answer. However, it is not clear that this would be true. If, as Quine says, familiarity of principle is one test of deciding among explanatory systems, then, even for us imperfect beings, it seems likely that at any one time, if several systems

are maximal, except for familiarity, then one of these will be most familiar and thereby maximal. And would not there be, for an all-perfect being, just one system that does consist, and always has consisted of eternally familiar principles?

12. Quine and Sellars as Foundationalists

I find the preceding theory of justification that supplements a foundational thesis with explanatory coherence to be quite attractive. Furthermore, I believe that, contrary to most opinions, it is somewhat plausible to interpret both Quine and Sellars as holding some form of this combined thesis, in spite of some things they say which at first glance seem to imply a rejection of foundationalism. For example, Quine says:

Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision.¹⁹

And Sellars states:

Our aim [is] to manipulate the three basic components of a world picture: (a) observed objects and events, (b) unobserved objects and events, and (c) nomological connections, so as to achieve a world picture with a maximum of "explanatory coherence." In this reshuffle, no item is sacred.²⁰

In spite of these quotations, however, Quine and Sellars can be foundationalists if, as is plausible, we take their main thesis to be that no particular set of statements is immune to refutation and so none is irrevocably at the foundation of empirical justification. That is, empirical justification does not have "the given" at its foundation. Sellars is opposed to observation statements being given and thus unalterably at the foundation, because he claims that, in the scientific millennium, singular theoretical statements will provide the most accurate and comprehensive picture of what there is. Thus he wants to be able to justify the ultimate rejection of all nontheoretical, observation statements. But none of this requires that he reject foundationalism. Indeed, his rejection of "the" given is not a rejection of foundationalism:

To reject the myth of the given is to not commit

¹⁹ W. V. O. Quine, *From a Logical Point of View* (New York 1961), p. 43.

²⁰ Wilfred Sellars, *Science, Perception, and Reality* (New York 1963), p. 356.

oneself to the idea that empirical knowledge as it is now constituted has no rock bottom level of observation predicates proper. It is to commit oneself rather to the idea where even if it does have a rock bottom level, it is *still* in principle replaceable by another conceptual framework in which these predicates do not, *strictly speaking*, occur. It is in this sense, and in this sense only, that I have rejected the dogma of given-ness with respect to observation predicates.²¹

Quine's views are amazingly similar. He seems to hold that observation statements are individually verifiable and are the tests of explanatory systems. He says that his view of the relationship of observation sentences to our knowledge of what is true, is very much the traditional one:

observation sentences are the repository of evidence for scientific hypotheses . . . Sentences higher up in theories have no empirical consequences they can call their own; they confront the tribunal of sensory evidence only in more or less inclusive aggregates. The observation sentence, situated at the sensory periphery of the body scientific, is the minimal verifiable aggregate; it has an empirical content all its own and wears it on its sleeve.²²

This surely seems to be a foundational thesis, indeed, one quite similar to the sort it is plausible to ascribe to Sellars. That is, both men seem to hold a foundational theory of justification that utilizes explanatory coherence but avoids "the" given. Yet, it might be objected, Quine and Sellars differ importantly, because Quine keeps observation statements at the foundation, but Sellars

wants to depose them. But there is no disagreement here, because, as Quine characterizes observation sentences, they include the very singular, categorical "theoretical" statements that Sellars wishes to place finally at the foundation.

13. Conclusion: A Foundational Theory With Explanatory Coherence

Is it reasonable to attribute this "explanatory-coherence" foundational theory of empirical justification to Quine and Sellars? And is this theory I attribute to them reasonable? I shall leave the first question for others to answer, preferably Quine and Sellars. Regarding the second, I shall make only three points here. First, this theory avoids the usual objections to foundational theories because it does not require an incorrigible given, does not restrict the sentences at the foundation to those solely about "the given," and does not limit inferential justification to deduction and induction by enumeration.²³ Second, if, as it seems plausible to expect, the notion of maximum explanatory coherence can be successfully clarified and the conditions for when a report is confirmed or disconfirmed by experience can be specified plausibly, then I find no damaging objections to the theory. Third, because of the first two points, and because in this paper we have seen some evidence that its main nonfoundational opponents are implausible, there is some reason to draw the tentative conclusion that this theory is a prime candidate for the most reasonable theory of empirical justification.²⁴

University of Pennsylvania

Received October 7, 1976

²¹ Wilfrid Sellars, *Philosophical Perspectives* (Springfield, Ill. 1967), p. 353.

²² W. V. O. Quine, *Ontological Relativity and Other Essays* (New York 1969), pp. 88-89.

²³ However, for an objection to foundational theories that: (1) statements at a foundation must have a probability of 1, and (2) only logical truths have that probability, see Lehrer, *op. cit.*, pp. 150-151. I have argued against (1) in "On Acceptability without Certainty," and against (2) in "On the Certainty of Reports about What is Given."

²⁴ I wish to thank Jaegwon Kim and Ernest Sosa for their very helpful comments.

IV. CREATURES OF FICTION

PETER VAN INWAGEN

I

SOME philosophers say there are things that do not exist. In saying this, they mean to assert more than the obvious truth that, on some occasions, sentences like "Mr. Pickwick does not exist" can be used as vehicles of true assertions: They mean to assert that there *are*, there really *are*, certain objects that have, among their other attributes (such as jollity and rotundity), the attribute of non-existence. Let us call such philosophers Meinongians and their doctrine Meinongianism.¹ One argument for Meinongianism proceeds by examples drawn from fiction, or so the Meinongian would say. A typical *anti*-Meinongian, however, would probably want to describe a typical application of this method as follows: "My Meinongian friend uttered 'Mr. Pickwick does not exist' assertively. Then he described what he had done in uttering these words as his having 'given an example of a non-existent object'." Our typical anti-Meinongian has an obvious reason for so describing the Meinongian's argument. For he is, of course, going to go on to say something like: "But his description of what he did was incorrect; for even if the sentence he uttered was or expressed a truth, its subject-term, 'Mr. Pickwick,' does not denote anything. Therefore, he did not, in uttering this sentence, succeed in giving an example of *anything*, much less of something non-existent."

So the Meinongian thinks that "Mr. Pickwick" is a name for something and that what it names is

non-existent.² The typical anti-Meinongian thinks that "Mr. Pickwick" is not a name for anything. It will be noticed that their positions are contraries, not contradictories. It would also be at least formally possible to maintain that "Mr. Pickwick" is a name for something and that what it names exists. In this paper, I wish to defend just this thesis. More generally, I shall defend the thesis that there are things I shall call "creatures of fiction," and that every single one of them exists.

I shall show that this thesis has certain advantages over both the Meinongian and what I have called the "typical anti-Meinongian" theories of the ontology of fiction. Its advantage over the Meinongian theory is this: Meinongianism *either* involves a bit of technical terminology that has never been given a satisfactory explanation, *or else* necessitates an abandonment of what are commonly called "the laws of logic." And the theory I shall present does not have this drawback. Consider the Meinongian's claim that there are things that don't exist. If I were asked to render this claim into the quantifier-variable idiom, I would write

$$(\exists x) \sim (\exists y)(x=y).$$

But the result of prefixing a tilde to this formula is a theorem of logic. Now the Meinongian will probably think my translation wrong-headed. I would expect him to say something like, "Either '(\exists x)' means 'there is an x' or it means 'there exists an x.' You can't have it both ways. Let's say it means the latter. Then you need to introduce a

¹ This is merely a convenient label. It is a nice question whether Meinong himself was a "Meinongian" in the present sense. Meinong would certainly say that the English sentence "There are things that do not exist" expressed a truth. But this truth it expressed would be about "*ideale Gegenstände*," ideal objects, such as Platonic forms and numbers. These subsist [*bestehen*] but do not exist [*existieren*], since there *are* such things, but they are not spatio-temporal objects. (On this matter, see note 5 to the present essay.) The proposition I mean to express by "There are things that do not exist" is, I think, that which Meinong expresses by the words, "There are objects of which it is true that there are no such objects" ["... *es gibt Gegenstände, von denen gilt, dass es dergleichen Gegenstände nicht gibt* ..."] ("Über Gegenstandstheorie," *Alexius Meinong Gesamtausgabe*, edited by Rudolf Haller and Rudolf Kindiger, in collaboration with Roderick M. Chisholm [Graz, 1969-1973] Vol. II, p. 490). But it is not clear how seriously he meant these words to be taken, since he prefaced them with "One who was fond of paradoxical modes of expression could ... say." On this topic, see R. M. Chisholm's very enlightening paper (from which most of the points made in this footnote are drawn), "Beyond Being and Nonbeing" in *New Readings in Philosophical Analysis*, edited by Herbert Feigl, Wilfrid Sellars, and Keith Lehrer (New York, 1972), pp. 15-22.

² There is, of course, a good deal more to Meinongianism than its treatment of the referents of names like "Mr. Pickwick" and "Hamlet." Even if the theory presented in the present paper, which treats *only* of names drawn from works of fiction, is true, it might yet be the case that the Meinongian is right in thinking that "the golden mountain," "Pegasus," "the round square," and so on, denote certain objects and that these objects do not exist.

new piece of notation for the former—say, (Ix) .
Then symbolize my general claim as

$$(Ix) \sim (\exists y)(x=y).$$

I must confess I do not understand the words I have put into the Meinongian's mouth. I do not see any important difference between "there is" and "there exists," and, therefore, I do not see how it is I am supposed to use " (Ix) " and " $(\exists x)$." Nor is this the end of the Meinongian's difficulties with the usual "laws of logic." For if the Meinongian is asked, "About your Mr. Pickwick—has he an even number of hairs on his head?" he will answer (Dickens having been noncommittal on this matter), "He neither has nor lacks the property of having an even number of hairs on his head; he is therefore what I call an *incomplete object*." And to say this, I would say, is to say that some instances of

$$(\forall x)(Fx \vee \sim Fx)$$

are false or express falsehoods, though this formula is a theorem of logic. Of course the Meinongian will again find my symbolic formula ambiguous. He will perhaps insist that I choose between two universal quantifiers, one corresponding to " $\sim (Ix) \sim$," and the other to " $\sim (\exists x) \sim$," and will claim that he assents to the invalidity only of " $\sim (Ix) \sim (Fx \vee \sim Fx)$." But this does not make the distinction between " $(\exists x)$ " and " (Ix) " any easier to understand.

These difficulties make the typical anti-Meinongian's position look very attractive by comparison. This philosopher will admit that sentences like "Mr. Pickwick does not exist" and "Mr. Pickwick is jolly" may, in certain contexts (correcting someone who takes *The Pickwick Papers* for history; giving a summary of the plot of that novel), be used by their utterers to express truths. But, he will insist, the utterer of such sentences does not (except in a Pickwickian sense) *refer to* anyone or anything when he utters them. Or, if our anti-Meinongian does not mind talking about such sentences "out of context" (this is how we represented him earlier), he may say that their subject-term does not denote anything, or that they are not subject to existential generalization.

Of course our anti-Meinongian must be able to explain how it is we are able to use a declarative sentence having "Mr. Pickwick" as its grammatical subject to express a truth when "Mr. Pickwick" is not a name for anything. Usually, when such an explanation is demanded, the anti-Meinongian will produce a *paraphrase* of the "Pickwick"-sentence; that is, he will devise a sentence he claims "really

says the same thing as" the "Pickwick"-sentence, in which "Mr. Pickwick" does not even *seem* to function as a name. For example, the anti-Meinongian might offer as a paraphrase of "Mr. Pickwick is jolly" the sentence "If *The Pickwick Papers* were not a novel but a true record of events, then there *would have been* a jolly man called 'Mr. Pickwick'." I do not say this would be a *good* paraphrase. Some philosophers might argue that it fails as a paraphrase because it contains a reference to a novel, which the original does not, and, moreover, contains a reference to the *words* "Mr. Pickwick," which the original does not. (Of course, it would be possible to reply that the original makes a *covert* reference to both these things, which the paraphrase has the virtue of making explicit.) Whether or not this is a good paraphrase, however, it is a good *example* of the sort of thing I mean by "paraphrase."

Well, what is wrong with "typical anti-Meinongianism"? Or perhaps we should ask, What is wrong with adopting it as a *program*?, since, at least as I have presented it, it is not an ontological doctrine, but rather a proposal to stick to a certain rather vaguely defined method in dealing with a certain class of ontological problems. I have no particular *a priori* objection to anti-Meinongianism, which has the advantages of leaving the laws of logic unviolated and of requiring no mysterious terminology. And perhaps it appeals to a "robust sense of reality," or, at least, appeals to those philosophers who like to describe themselves as having such a sense. But I am not sure the program it proposes can be carried out. I have no knock-down argument for this, though I shall in the sequel display some sentences that, at the very least, are going to force the typical anti-Meinongian to produce very messy paraphrases. But I am not, in this paper, primarily interested in refuting either "typical anti-Meinongianism" or Meinongianism. I wish only to present, in a very sketchy way, the *prima facie* cases against the way these theories deal with those of our assertions and beliefs that are "about fictional entities," in order to present a *prima facie* case for the rival theory I shall presently propose.

II

At this point I am reluctantly going to abandon Mr. Pickwick, despite the fact that, like that old war horse, Pegasus, he has a good claim to tenure in the office of Exemplary Nonentity. I turn

instead to Mrs. Gamp. I do this because I have been able to find sentences "about her" that serve my purposes better than any sentences I know of "about Pickwick." Here are three of them:

- (1) She was a fat old woman, this Mrs. Gamp, with a husky voice and a moist eye, which she had a remarkable power of turning up, and only showing the white of it (*Martin Chuzzlewit*, XIX)
- (2) Mrs. Sarah Gamp was, four-and-twenty years ago, a fair representation of the hired attendant on the poor in sickness (From Dickens's preface to an 1867 edition of *Martin Chuzzlewit*)
- (3) Mrs. Gamp . . . is the most fully developed of the masculine anti-women visible in all Dickens's novels (Sylvia Bank Manning, *Dickens as Satirist* [New Haven, 1971] p. 79).

Now a very naive Meinongian might describe what the authors of these sentences were doing in writing them as follows. "There is a certain non-existent woman, Mrs. Gamp, and both Dickens and Professor Manning wrote about her. In writing (2), Dickens asserted that, in 1843, she was a fair representation of the hired attendant on the poor in sickness; in writing (1), he asserted of her that she was [in 1843?] fat, old, husky-voiced, and so on, while, in writing (3), Professor Manning asserted of her that she was [in 1971?] the most fully developed of the masculine anti-women visible in all Dickens's novels."

Now whatever else may be wrong with the naive Meinongian's description of what the writers of these sentences were doing, surely he is wrong in so assimilating the writing of (1) to the writing of (2) and (3). Sentences (2) and (3) were used by their authors as the vehicles of assertions; (1) was not. It would make sense to say that the authors of (2) and (3), in writing these sentences, wrote something true or wrote something false. If the average "hired attendant on the poor in sickness" in 1843 was rather like the popular picture of Florence Nightingale, then Dickens, in writing (2), wrote something *false* in the same sense as that in which he would have written something false if he had

written that he had composed *Martin Chuzzlewit* while living in China. But if someone had been looking over Dickens's shoulder when Dickens was writing (1), and had said to him, "No, no, you've got her all wrong. She is quite thin, about twenty-four, and her voice is melodious," this would simply have made no sense.

Thus, there is a certain sense in which the fact that novelists do things like writing sentence (1) is not directly relevant to questions about the ontology of fictional entities. There is no point in debating what sort of thing Dickens was writing about when he wrote (1) or debating what sort of fact or proposition he was asserting, since he was not writing about anything and was asserting nothing.³ Sentence (1) does not represent an attempt at reference or description.⁴

Sentences (2) and (3), however, do represent assertions: their authors in writing them are expressing propositions of some sort, and these propositions *seem* to be about Mrs. Gamp. A more sophisticated Meinongian will say that *these* sentences, at any rate, are about a non-existent entity, even if (1) is not. Like this Meinongian, I suggest we take Dickens's and Manning's uses of sentences (2) and (3) at face value: as assertions about a certain entity called "Mrs. Gamp." Unlike him, however, I suggest that what Dickens's and Manning's assertions are about exists. In Part III, I shall attempt to answer certain questions that must be answered by anyone who takes these suggestions seriously.

III

Question: Why should we take (2) and (3) as being *about* Mrs. Gamp, in any sense except the uninteresting sense in which "The average American owns 1.02 cars" is "about the average American"? *Answer:* Because, while "the average American" is not a name for anything, "Mrs. Gamp" is a name for something—or, if you prefer, there is such a thing as Mrs. Gamp—and if there is such a thing as Mrs. Gamp, then *obviously* (2) and (3) are about her. *Question:* But why do you say there is such a thing as Mrs. Gamp? *Answer:* Because there *are* such things as characters in novels. And if there are such things as characters in novels,

³ This is a very important point. The reader who does not concede it will get very little out of reading further. The argument in the text of the present paper is not sufficient to establish it. For arguments that are sufficient, see Alvin Plantinga, *The Nature of Necessity* (Oxford, 1974) Ch. VIII, pp. 153–163 especially, and J. O. Urmson, "Fiction," *American Philosophical Quarterly*, vol. 13 (1976), pp. 153–157.

⁴ This is true despite the fact that (1) is what is called a "descriptive" sentence by literary critics. This term is applied to sentences of fiction that would be descriptive sentences in a literal sense if the works of fiction in which they occurred were historical records.

then Mrs. Gamp is one of them. Anyone who said there were such things as characters in novels, and went on to say that there was no such thing as Mrs. Gamp would simply be factually ignorant. He would be like someone who said that there were such things as irrational numbers, but no such thing as π . *Question*: But why do you say there are such things as characters in novels? *Answer*: Consider sentences like the following:

- (4) There are characters in some 19th-century novels who are presented with a greater wealth of physical detail than is any character in any 18th-century novel
- (5) Some characters in novels are closely modeled on actual people, while others are wholly products of the literary imagination, and it is usually impossible to tell which characters fall into which of these categories by textual analysis alone
- (6) Since 19th-century English novelists were, for the most part, conventional Englishmen, we might expect most novels of the period to contain stereotyped comic Frenchmen or Italians; but very few such characters exist.

These sentences, or the propositions expressed by them, certainly *seem* to assert that there are things of a certain sort: if anyone were to utter one of these sentences assertively, it would *seem* that what he would say could be true only if there were such things as characters in novels. Take (4), for instance. If I were asked to render this sentence into the quantifier-variable idiom, I would produce something like:

- (4)* $(\exists x)(x \text{ is a character in a 19th-century novel} \ \& \ (\forall y) [y \text{ is a character in an 18th-century novel} \supset x \text{ is presented with a greater wealth of physical detail than is } y])$.

And, by the rules of formal logic, (4)* yields:

- (7) $(\exists x)(x \text{ is a character in a 19th-century novel})$.

Therefore, since (4) is true, or expresses a true proposition, or is such that if it were uttered in appropriate circumstances its utterer would say something true, and since (4)* is a correct translation of (4) into the regimented idiom of formal logic, and since the rules of formal logic are truth-preserving, and since "There are such things as characters in 19th-century novels" is a correct translation of (7)

into ordinary English, it follows that there are such things as characters in 19th-century novels. And from this it follows that there are such things as characters in novels.

Question: There are several premises in that argument that one might want to look at rather carefully. But let's grant for the sake of argument that there are such things as characters in novels; What do you mean by saying they *exist*? *Answer*: Just what you granted and no more. "There is no cure for cancer," "There is no such thing as a cure for cancer," "A cure for cancer does not exist," and "There exists no cure for cancer" all mean more or less the same thing. Some of them might be slightly more appropriate things to say in certain contexts of utterance than others, but it is not possible that anyone should utter one of these sentences and thereby say something true, and someone else simultaneously utter another of them and thereby say something false. And the same point applies whether we are talking about cures, cabbages, countries, cylindrical algebras, or characters in novels. There is (or exists) a tendency in some quarters to think that "there are" is a harmless and rather empty expression, while "exist" is such an important word that anyone who uses it takes on a great weight of ontological responsibility. But "exist" in (6) above, and "there are" in (4), and, for that matter, "some" in (5), have just exactly the same sort of import. And this is *always* the case, no matter what sort of thing we are talking about.⁵

Question: But if Mrs. Gamp really existed, couldn't I, if I had been alive in 1843, have gone and talked with her? *Answer*: Obviously not, since she did exist then (she exists *now*, in fact) and you couldn't have. But now you are touching on matters I will deal with in Part IV. *Question*: But I find her mysterious. What sort of thing is she? What ontological categories does she belong to? *Answer*: Well, she is, as I said, a character in a novel. And characters in novels are members of a category of things I shall call *creatures of fiction*. Some things belonging to this category that are not characters in novels are the Wife of Bath, Polyphemus, the Forest of Arden, Dotheboys Hall, and Professor Moriarty's book *The Dynamics of an Asteroid*. And creatures of fiction belong to a broader category of things I shall call *theoretical entities of literary criticism*, a category that also includes plots, sub-

⁵ "He [Rush Rhees] thinks we use 'exists' 'chiefly in connexion with physical objects' . . . The nearest newspaper shows the contrary; 'conditions for a durable agreement do not yet exist', or the like, is the commonest currency of journalism." P. T. Geach, review of Rush Rhees's *Without Answers* in *The Journal of Philosophy*, vol. 68 (1971), pp. 531-532.

plots, novels (as opposed to tangible *copies* of novels), poems, meters, rhyme schemes, borrowings, influences, digressions, episodes, recurrent patterns of imagery, and literary forms ("the novel," "the sonnet"). Or this category includes such things as these if there *are* any such things as these. (Cf. the sequence: the thing that caused this trace in the cloud-chamber, electron, sub-atomic particle, theoretical entity of physics.)

Question: And just what are theoretical entities?

Answer: That's a very good and not very clear question. Of course, it's not *your* fault it's not clear. But let's look at it this way. There are various "theoretical disciplines" like physics and literary criticism. (The former of these is a science, the latter not.) There are various sentences in which the "conceptual machinery" of these disciplines is inextricably embedded: you couldn't *say* what these sentences say without employing theoretical vocabularies like the ones these disciplines employ. Some of these disciplines may be such that we are comfortable with saying that nothing interesting that can be said only by means of their special vocabularies is ever *true*. (Astrology, for example.) But for many disciplines (physics, say) this would be an absurd thing to say. (Like many absurd things, it has been said.) And I think it would be absurd to think that nothing that can be said only in the language of literary criticism is true, especially if we take "literary criticism" to include all "informed" discourse about the nature, content, and value of literary works. And, sometimes, if what is said in a piece of literary criticism is to be true, then there must be entities of a certain type, entities that are never the subjects of non-literary discourse, and which make up the extensions of the theoretical general terms of literary criticism. It is these that I call "theoretical entities of literary criticism." To say this much, however, is not to answer the question, Which theoretical terms of criticism must be taken as having special sorts of entity as their extension?, or the question, Which, if any, of these terms is in principle eliminable from critical discourse? But I think that "discourse about characters," which is the sort of critical discourse that is our present concern, is not *easily* eliminable from literary criticism. I see no way to do it, at any rate. That is, I see no way to paraphrase sentences (4), (5), and (6), and others like them, in such a way as to produce sentences that seem to "say the same thing" and which do not involve "quantification over creatures of fiction." (This is not a very carefully stated claim; taken

literally, it is obviously false, since it would be easy enough to paraphrase these sentences in such a way that the paraphrases involve quantification only over, say, unit sets whose members are creatures of fiction. Don't take it *that* literally.)

Question: But isn't finding such a paraphrase a fairly easy task? Take sentence (4). It seems obvious that, once all the 18th- and 19th-century novels have been written, once certain novelists, writing during certain centuries, have finished putting words down on paper in a certain order, then the *facts* relevant to the truth-value of (4) are completely settled, and the truth-value of (4) is determined. Suppose a philosopher is guided by this thought and suppose he is willing to admit novels and classes into his ontology but not willing to admit creatures of fiction. Then he might simply invent an open sentence, e.g., "*x* dwelphs *y*," which is satisfied only by pairs of classes of novels, and which (however its semantics is to be spelled out in general) is satisfied by the pair <the class of 19th-century novels, the class of 18th-century novels> if and only if what is expressed by (4) is *true*. Couldn't this philosopher then offer "The class of 19th-century novels dwelphs the class of 18th-century novels" as a paraphrase of (4) that does not involve quantification over creatures of fiction? *Answer:* In a certain trivial sense, this device works. The "dwelphs"-sentence, on the usage of "dwelph" your philosopher has stipulated, expresses the same proposition as (4)—or, at least, expresses a proposition necessarily equivalent to that proposition. But this is not *all* that is required of a good paraphrase. To see this, consider

- (8) Every female character in any 18th-century novel is such that there is some character in some 19th-century novel who is presented with a greater wealth of physical detail than she is.

Now the proposition expressed by this sentence is certainly a logical consequence of the proposition expressed by (4). I, who accept the existence of things I call "characters in novels," can account for this fact: (8) is a translation into ordinary English of a sentence in the regimented quantifier-variable idiom that is a *formal* consequence of (4)*, which is a translation of (4) into the quantifier-variable idiom. But if someone were to paraphrase (8) into a sentence consisting of two class-terms flanking a relation-sign (this is what you just now imagined a philosopher doing with (4); let's imagine the paraphrase of (8) looks like this:

"The class of 18th-century novels praphs the class of 19th-century novels"), he would have no way to *account for* the fact that his paraphrase of (8) expresses a proposition that is a logical consequence of the proposition expressed on his usage by "The class of 19th-century novels dwelphs the class of 18th-century novels."

Because of this consideration, I lay down the following condition of adequacy on any attempt to paraphrase away quantification (or apparent quantification) over "creatures of fiction": an adequate paraphrase must not be such as to leave us without an account of the logical consequences of (the propositions expressed by) the paraphrased sentences. Almost certainly, any paraphrase that satisfies this condition will have a quantificational structure not much simpler than the (apparent) quantificational structure of its "original."

I am not saying it would be *impossible* to devise paraphrases of (4), (5), (6), and similar sentences, that satisfy this condition. Probably the most promising candidates for such paraphrases would involve quantifications mainly over "character-names" (e.g., "Sophia Western," "Rodya Raskolnikov"), the sentences of fiction in which these "names" occur, and the open sentences that can be got by replacing them in these sentences by variables. I do not myself see any way of doing this, but perhaps someone cleverer than I will think of some way to do it.

But (if I may ask *you* a question) why should anyone bother to try to construct such paraphrases? It would probably be very difficult to do this, and the paraphrases would probably be long and messy if they could be got at all; and maybe they *couldn't* be got, in which case one would have been wasting one's no doubt valuable time in trying to get them. So why embark on such an enterprise?

Compare this question with a similar question in the ontology of mathematics: Why should anyone bother to try to construct paraphrases of such sentences as "There is a prime number between 18 and 20" and "There exists a least number that can be expressed in more than one way as the sum of two cubes" that do not involve quantification over numbers? One reason someone might have (there could be others, such as a desire to indulge one's "taste for desert landscapes") is that he finds the idea of there being such things as numbers a very puzzling one. How *could* there be things (he might ask) that *exist* in the same sense as that in which you and I exist, which *have properties* in the same sense as that in which "moderate-sized speci-

mens of dry goods" have properties, and *bear relations* to one another in the same sense as that in which stones and bits of stick bear relations to one another, and which are, nevertheless, intangible and eternal?

And perhaps there are philosophers who find the very idea of creatures of fiction objectionable. We shall investigate just what it is that might be objectionable about them in Part IV. (That is, we shall investigate the question, What is it that is objectionable about creatures of fiction *per se*? We shall not attempt to answer objections that stem from a conviction that theoretical or abstract entities *in general* are objectionable.)

IV

Just what is puzzling or objectionable about Mrs. Gamp? Well, consider the following properties

- being old
- being fat
- being fond of gin
- being named "Sarah Gamp"
- having a friend called "Mrs. Prig."

No one (in 1843) had *all* these properties. (If by chance someone did, we could enlarge the list.) But, someone may argue, Mrs. Gamp has, or had, or is supposed to have had all these properties. Thus, Mrs. Gamp is a very puzzling entity indeed, and, since this point could be generalized so as to apply to all "creatures of fiction," any ontology that includes them is objectionable.

This is a very powerful argument. The Meinongian and I are equally obliged to respond to it. *He* will deny the premise that no one has (or had in 1843) these properties. To the objection that it is simply an empirical fact that no one in 1843 had these properties, he will reply that this is not an empirical fact. What is an empirical fact, he will say, is that no one *existent* in 1843 had these properties. This reply is not available to me, however, who regard "no one existent" as a long-winded way of saying "no one." I shall deny the premise that Mrs. Gamp has or had these properties. This thesis will probably strike the reader as odd. But I think it is not so odd as it sounds, and, in fact something formally very much like it is a familiar philosophical doctrine.

Let us turn for a moment to the Cartesian psychology, that is, to the doctrine that each of us is an immaterial substance. Suppose someone were to say to Descartes, "Your theory is obviously wrong since Jones here has all the following properties:

being tangible
 being six foot tall
 weighing 190 pounds
 being mainly pinkish-white in color,

and these properties could not be had by an immaterial substance." Descartes, of course, will reply that Jones does *not* have the properties on this list. What he *has*, Descartes tells us, are properties like these:

thinking about Vienna
 being free from pain
 being in a state of Grace
 animating a body.

Nonetheless, Descartes will concede, Jones bears a certain intimate relation to the properties on the former list that is not the relation of "having" or "exemplifying" but, rather, the relation of "animating a body that has or exemplifies." Descartes will further concede that in ordinary speech we often say "is" when *strictly speaking* we should say "animates a body that is": what looks like predication in ordinary speech is not always predication.

I want to say something similar with respect to Mrs. Gamp. Here are some of the properties she *has*:

being a character in a novel
 being a theoretical entity of literary criticism
 having been created by Dickens
 being a satiric villainess.

Moreover, if Dickens and Professor Manning are right, she has the following properties:

being a fair representation of the hired attendant on the poor in sickness in 1843
 being the most fully developed of the masculine anti-women visible in all Dickens's novels.

Moreover, *if* existence is a property (an hypothesis the correctness of which I take no stand on in this paper), she has that, too. Thus, the properties Mrs. Gamp *has* are just those "literary" properties that are appropriate to what she is: a theoretical entity of criticism. If she shares any properties with you and me, they are "high-category" properties like existence, self-identity, and non-identity with any ordinal number.

Now just as Jones, according to Descartes, does not *have* the property of being tangible, so Mrs. Gamp, according to me, does not *have* the property of being fat. Nevertheless, just as Jones, on the Cartesian view, bears a certain intimate relation

to tangibility, Mrs. Gamp bears a certain intimate relation to fatness. And just as, on the Cartesian view, we may say "Jones is six foot tall" and be talking *about* an immaterial substance without thereby *predicating* being six foot tall of that immaterial substance, so, on the present view, we may say "Mrs. Gamp is fond of gin" and be talking *about* a theoretical entity of criticism without thereby *predicating* fondness for gin of that theoretical entity of criticism.

Such circumlocutions as were employed in the preceding paragraph are cumbersome. Clearly it would be to our advantage to introduce some name or other for this special relation that Mrs. Gamp bears to fatness and fondness for gin. For want of a better name, I shall call it "ascription." Thus, we might write, "Fatness is ascribed to Mrs. Gamp" if we wished to express the proposition expressed in ordinary usage by "Mrs. Gamp is fat" without using what looks like the "is" of predication. But I think we had better make ascription a three-term relation. For consider the fact that the sentences "Shiela Smith was fat" and "Shiela Smith was not fat" might appear in one and the same novel, owing either to the author's inadvertence or to the passage of narrative time. Suppose there was such a novel. Should we, describing its content in a review, write "Shiela Smith was fat"? This would be misleading. If the two sentences appeared in the novel owing to the author's inadvertence, clearly we ought to write something like "Shiela Smith is described as fat in Chapter Four. But in Chapter Six we are told she is not fat." In order to provide for cases like this, let us use the following open sentence in asserting that the ascription relation holds: " x is ascribed to y in z " [hereinafter abbreviated " $A(x, y, z)$ "]. For any x , y , and z , if $A(x, y, z)$, then x is a property, y is a creature of fiction, and z is what I shall call a "place." A *place* is either a work of fiction (such as a novel, short story, or narrative poem) or a part or section thereof, even a part or section that is so short as to be conterminous with a single (occurrence of a) sentence or clause. The proposition commonly expressed by "Mrs. Gamp is fat" we may express by " $A(\text{fatness, Mrs. Gamp, Martin Chuzzlewit})$ " or " $A(\text{fatness, Mrs. Gamp, Ch. XIX of Martin Chuzzlewit})$ " or " $A(\text{fatness, Mrs. Gamp, the only occurrence of (1) in Martin Chuzzlewit.})$ " Or, if these assertions be thought to be too definite to capture the normal sense of "Mrs. Gamp is fat," we could write

$(\exists x)A(\text{fatness, Mrs. Gamp, } x),$

or even the conjunction of this last sentence with

$\sim (\exists x)A(\text{non-fatness, Mrs. Gamp, } x).$

Now the word "ascription" is not a good word for this relation. It is misleading. But I cannot think of a word that would not be misleading. In order to see how this term could mislead, consider the following cases. Suppose Dr. Leavis should write an essay called "Current Nonsense about Dickens," and suppose this essay contained the sentence (used in a straightforward way as the vehicle of an assertion), "Mrs. Gamp is thin." Given the ordinary meaning of "ascription," someone might be led to describe this state of affairs by writing

(9) $A(\text{thinness, Mrs. Gamp, "Current Nonsense about Dickens"})$.

But this would be a mistake. "Current Nonsense about Dickens" is not a fictional work but a critical essay, and hence is not a "place" and hence sentence (9) does not express a truth.

Or consider the following sentence

(10) $A(\text{vanity, Napoleon, } War \text{ and } Peace).$

Although in the ordinary sense of "ascribe" it is true that Tolstoy ascribed vanity to Napoleon in *War and Peace*, nonetheless (10) is not true since Napoleon is not a creature of fiction.⁶

Finally, it is important to realize that it does not follow from the truth of " $A(\text{fatness, Mrs. Gamp, } Martin \text{ Chuzzlewit})$ " that Dickens made a certain sort of silly mistake: viz., that in writing sentence (1), he (mistakenly) ascribed to a certain theoretical entity of criticism the property of being fat. For it is not the case that when Dickens wrote (1) he was ascribing any property to anything, either in our technical sense of "ascribe" or in the ordinary sense; to ascribe a property to something (in either sense of "ascribe"), Dickens would have had to have been expressing some proposition when he wrote (1), and, as I have said, he did not express any proposition by writing (1).

I shall not attempt to give a definition of " $A(x, y, z)$." This three-place predicate must be taken as primitive. But I think its sense is fairly easily grasped. We are saying something *true* about the relations that hold between the novel *Martin Chuzzlewit*, the main satiric villainess of that novel, and the property fatness when we say, "Mrs.

Gamp, a character in *Martin Chuzzlewit*, is fat." And I think that we are *not* saying that the relation of *exemplification* holds between Mrs. Gamp and fatness when we say this, since that would *not* be true: if anything *exemplifies* fatness, then it occupies a certain region of space-time and you and I (if we are appropriately located in space-time) can touch it; but Mrs. Gamp is a theoretical entity of criticism, and we could no more touch her than we could touch a plot or a sonnet. " $A(\text{fatness, Mrs. Gamp, } Martin \text{ Chuzzlewit})$ " is nothing more than the way I choose to express what we normally express by the above sentence, and that is all the explanation I am able to give of the use of " $A(x, y, z)$." In order to see the difficulties one encounters in attempting to provide an explicit definition of " $A(x, y, z)$ " let us look at two rather obvious proposals.

First, one might say that " $A(\text{fatness, Mrs. Gamp, } Martin \text{ Chuzzlewit})$ " simply means that if there were a real woman like Mrs. Gamp, that woman would be fat. But this would be either wrong or circular. A "real" woman (I suppose that means, strictly, a woman) could not be "like" Mrs. Gamp, if that means having the properties Mrs. Gamp *has*, since no "real" woman could be a character in a novel, be a theoretical entity of criticism, or have been created by Dickens. Of course, a "real" woman could be "like" Mrs. Gamp in *having* the properties that are *ascribed* to Mrs. Gamp in *Martin Chuzzlewit*. But, obviously, "like" in this sense cannot occur in a *definition* of the ascription relation.

Or one might want to say that " $A(\text{fatness, Mrs. Gamp, } Martin \text{ Chuzzlewit})$ " means that if *Martin Chuzzlewit* were not a novel but a true record of events, then there *would be* a woman called "Mrs. Gamp" and she would be fat. But there is a subtle difficulty hidden in this proposal. Consider " $A(\text{fatness, the main satiric villainess of } Martin \text{ Chuzzlewit, } Martin \text{ Chuzzlewit})$ " or " $A(\text{fatness, the most important character in } Martin \text{ Chuzzlewit introduced in Ch. XIX, } Martin \text{ Chuzzlewit})$." If we were to try to understand these sentences (which express truths) in the way proposed, we should come up with "If *Martin Chuzzlewit* were not a novel but a true record of events, then there would be a woman denoted by 'the main satiric villainess of *Martin Chuzzlewit*' and she would be fat" in the case of

⁶ That is, normally "Napoleon" is used to denote a certain *man*. It may be that when critics discuss *War and Peace*, they at least sometimes use "Napoleon" to designate a certain creature of fiction that is (of course) numerically distinct from the man Napoleon. (See n. 8.) If this is the case, then the triple $\langle \text{vanity, the creature of fiction sometimes called "Napoleon" by critics discussing } War \text{ and } Peace, War \text{ and } Peace \rangle$ satisfies " $A(x, y, z)$." But $\langle \text{vanity, the man Napoleon, } War \text{ and } Peace \rangle$ does not satisfy " $A(x, y, z)$."

the first of these sentences, and a similar piece of nonsense in the case of the second.

Let us see why this difficulty arises. How it is we are able to use the *proper name* "Mrs. Gamp" to refer to a certain creature of fiction? Normally, an object gets a proper name by being dubbed or baptized. But no one ever dubbed or baptized the main satiric villainess of *Martin Chuzzlewit* "Mrs. Gamp."⁷ There is no corresponding problem about how it is this creature of fiction is denoted by "the main satiric villainess of *Martin Chuzzlewit*," for this is a quite straightforward definite description that names what we also call "Mrs. Gamp" for the same reason that "the tallest structure in Paris in 1905" names what we also call "the Eiffel Tower": in each of these cases, a definite description denotes a certain object in virtue of a certain property that that object has uniquely. I think that if we are to have a satisfactory theory of how it is that we manage to refer to particular creatures of fiction, this theory will have to treat such *descriptions* as "the main satiric villainess" as the *primary* means of reference to these objects, and *proper names* as a secondary (though more common) means of reference. I would suggest that the *only* reason "Mrs. Gamp" denotes a certain creature of fiction is that that creature of fiction satisfies the open sentence "A (being named 'Mrs. Gamp,' *x*, *Martin Chuzzlewit*)."⁸ Thus, any such explanation of "A(*x*, *y*, *z*)" as the one we are considering must be circular, for the fact that a certain creature of fiction (such as the main satiric villainess of *Martin Chuzzlewit*) is denoted by a "fictional proper name" (like "Mrs. Gamp") can be explained only in terms of the ascription relation. Moreover, even if such an explanation were not circular, it would not show us how to understand instances of "A(*x*, *y*, *z*)" in which the '*y*'-position is filled by a definite description involving a uniquely exemplified "literary" property, such as "the character introduced in Ch. XIX."

⁷ Of course, Dickens might have said to himself, "I'll call the character I'm about to introduce, 'Sarah Gamp.'" He might even have said this out loud, to an audience. And he might *not* have done (and probably *did* not do) these things. Whether he did or not, Mrs. Gamp would be called "Mrs. Gamp." Therefore, such an utterance on Dickens's part (if it occurred) was not an act of dubbing, and had nothing in particular to do with the fact that it is proper for us to call the important female character introduced in Ch. XIX of *Martin Chuzzlewit*, "Mrs. Gamp." Moreover, it is obvious that our ability to refer to this character as "Mrs. Gamp" does not derive from some early Dickens critic's having said, "I hereby dub the main satiric villainess of *Martin Chuzzlewit*, 'Sarah Gamp.'"

⁸ That is, we have embodied in our rules for talking about fiction a convention that says that a creature of fiction *may* be referred to by what is (loosely speaking) "the name it has in the story." It is owing to a similar convention that we use *personal* pronouns in connection with entities that are literally *not* persons: we call Mrs. Gamp "she" because, though she is not a woman, the property of being a woman is ascribed to her. (Cf. the Cartesian's explanation of how it is we are able to use demonstrative pronouns and adjectives in referring to substances that are *literally* non-spatial.) Once we have grasped the ascription relation, it is easy enough to explain and analyze our use of "mixed" descriptions like "the fat old nurse introduced in Ch. XIX." (Cf. the Cartesian's explanation of our use of "the fattest person thinking of Vienna.")

It is because of these difficulties that I am content to take "A(*x*, *y*, *z*)" as primitive. (But one small stipulation: let us say that a property *entails* a second property if it is impossible that something have the former and not have the latter; then ascription is "closed" under entailment. That is, $(\forall x)(\forall y)(\forall z)(\forall w)[A(x, y, z) \ \& \ x \text{ entails } w. \supset A(w, y, z)]$.)

This, in essence, is the theory of creatures of fiction I want to present. To recapitulate:

- (a) Sentences like (2), (3), (4), (5), and (6) are about characters in novels. Characters in novels belong to a class of entities I call "creatures of fiction," which class is a subclass of the class of theoretical entities of criticism.
- (b) Sentences like (1), that is, typical narrative or descriptive sentences taken from works of fiction, are not about creatures of fiction. They are not about anything. They are not used by their authors as the vehicles of assertions.
- (c) Creatures of fiction exist and obey the laws of logic, just as everything else does.
- (d) Creatures of fiction *have* or *exemplify* only "literary" properties such as *being introduced in Ch. XIX* and "high-category" properties like self-identity. They do not have such properties as being human or being fat, despite the fact that ordinary-language sentences like "Mrs. Gamp is fat" can be used to express true propositions about them. (We call the relation that, I claim, is asserted by sentences like this to hold between creatures of fiction and such everyday properties as you or I might have, "ascription.")

I shall conclude by showing how three problems about fictional entities can be easily solved on the present theory.

V

The first problem: How should we deal with questions like "Has Mrs. Gamp an even number of hairs on her head?" The Meinongian, remember, says that Mrs. Gamp neither has nor lacks the property of having an even number of hairs on her head. The typical anti-Meinongian says (roughly) that "she" is not "there" either to have or to lack such a property. I say she simply lacks it, just as she lacks the properties of being material, being human, having hair, and being bald. Moreover, for any property whatever, Mrs. Gamp either has that property or lacks it.

But what is *not* true is this:

$(\forall x)(\exists y)[A(x, \text{Mrs. Gamp}, y) \dot{\vee} A(\text{not having } x, \text{Mrs. Gamp}, y)]$.

(Here ' $\dot{\vee}$ ' represents exclusive disjunction.) And, in particular, *this* is not true:

$A(\text{having an even number of hairs on one's head, Mrs. Gamp, Martin Chuzzlewit}) \dot{\vee} A(\text{not having an even number of hairs on one's head, Mrs. Gamp, Martin Chuzzlewit})$.⁹

This disjunction is false simply because Dickens was noncommittal about the oddness or evenness of the number of Mrs. Gamp's hairs. It is these facts, I think, that the Meinongian perceives through a glass darkly when he says that a creature of fiction is an incomplete object.

The second problem: Mrs. Gamp appears to have incompatible properties. For consider the properties

being a woman

having been created by Dickens.

Now, since Dickens was not God, it would seem that nothing could have both these properties. To this the Meinongian will reply that I am mistaken in thinking that only God can create things that fall within the extension of the property *being a woman*; what only God can do, the Meinongian will tell us, is to create things that fall within this extension and which are also *existent*. Dickens is perfectly

capable of creating women (the Meinongian holds); he is, however, incapable of creating *existent* women. (Or perhaps, the Meinongian will want to say that Dickens did not *create* Mrs. Gamp, but did something more like *discovering* her in the realm of *Sosein*. I shall not explore the ramifications of this thesis, except to point out that it makes the creativity of the novelist seem very like the "creativity" of the flower-arranger.) This reply seems to me to be unsatisfactory. What reason do we have for saying that Mrs. Gamp has the property *being a woman*? Only that she "has it in the story." Well, "in the story" she also has the property of existence, at least if existence *is* a property (but this the Meinongian assumes).¹⁰ Therefore, I submit, if the Meinongian has any good reason for saying that the extension of *being a woman* includes Mrs. Gamp, he has an equally good reason for saying that the extension of *being existent* includes her.

This problem is, of course, trivial on the theory I am proposing: Mrs. Gamp *has* only the second of these properties; the first is not *exemplified* by her, but is, rather, ascribed to her.

The third problem: Consider the definite descriptions, "the main satiric villainess of *Martin Chuzzlewit*" and "the character in *Martin Chuzzlewit* who appears in every chapter." Since no character in *Martin Chuzzlewit* appears in every chapter, there seems to be an important difference between these two descriptions. I can describe this difference very simply. The first of these descriptions denotes something (Mrs. Gamp), the second does not: there *is* no character in *Martin Chuzzlewit* who appears in every chapter; no such character *exists*. The Meinongian will want to say that each of these descriptions denotes something, and that what each denotes fails to exist. But there is obviously an important ontological difference between the ways in which these two descriptions relate to the world. I leave it to the Meinongian to explain how this difference is to be spelled out in his terms.¹¹

Syracuse University

Received August 25, 1976

⁹ But the proposition that $A(\text{having or not having an even number of hairs on one's head, Mrs. Gamp, Martin Chuzzlewit})$ is true, owing to the fact that ascription is closed under entailment.

¹⁰ Note that "in the story" Mrs. Gamp's imaginary friend Mrs. Harris, unlike Mrs. Gamp herself, does *not* exist.

¹¹ A little reflection on this problem should show that what I call creatures of fiction cannot be or be among the "merely possible individuals" the existence of which is assumed in most informal explanations of Kripke's and others' systems of formal semantics for quantified modal logic. Creatures of fiction could not serve as merely possible objects because, for one thing, they are *actual* objects. But then what of the sentence with which we began, "Mr. Pickwick does not exist," and our concession that it could be used to express a truth? This is a very complicated question. Part of the answer is this. The utterer of such a sentence would probably be addressing himself to someone who had mistaken discourse about Pickwick for discourse about a *man*, owing to the fact that the ascription relation is expressed in English by what sounds like the apparatus of predication. He would probably be expressing the proposition that there is no such *man* as Pickwick, or, more precisely, the proposition that nothing *has* all the properties *ascribed* to Pickwick.

V. THE NEGLECT OF RELATIVE PREDICATES IN MODERN PHILOSOPHY

"Logic is the backbone of philosophy."

May Sinclair

CHARLES HARTSHORNE

I. DEPENDENCE AND INDEPENDENCE

THE story I shall try to tell has not as a whole been told before. Some parts of it can be found in my writings. Most of the key historical facts are well presented in a book by Passmore.¹ In what follows, numbers in parentheses will identify pages in this work. (Where page numbers differ in hard-back and Penguin paperback forms, hard-back page numbers will occur first and be larger. The content is the same in those parts to which I shall refer.) In my long philosophical development I have reacted with some care to many of the philosophers Passmore expounds, and I acknowledge the almost un-failing fidelity and adroitness of his expositions. But collectively they have for me implications that do not occur to him—or, it seems, to other writers of works of reference.² Reading Passmore, or authors he reports on, is not needed for understanding the "process" philosophy embodied in my account,³ but it should be helpful for judging how far the account is historical, rather than a series of straw men. And the historical aspect has its own interest.

A more distinctive label for "relative" predicates might be "relativizing" ones, those implying existential relativity, or dependence of actualities upon other actualities. Mere predicates of comparison are not here in question. I could not be taller than some contemporary on a remote planet were there no such contemporary, but this is no ground for denying that I could, in that case, have existed with my present height. Such predicates do not relativize the actualities compared, though they may relativize experiences of the compared terms to the terms.

The monistic thesis that all relations relativize their terms—as Green, e.g., held (pp. 57–58)—was with reason rejected by Moore, Russell, and my teacher R. B. Perry. But the contrary extreme, no relations relativize their terms, is not necessarily truer. If every entity depends on every other with which it is "in relation," then in no distinctive sense are there relativizing properties. Likewise, if no entity depends on any other, there is no point in talk about dependence—or, for that matter, ("principle of contrast") independence. Over-generalizing either concept destroys both. Monists and pluralists alike seem to miss this point.

Among important relativizing predicates are: X is influenced or caused by . . . , X temporally follows, X perceives or remembers. That these relativize X has often been denied or misinterpreted, e.g., by Hume. Some nonrelative predicates are: X influences or causally conditions . . . , X precedes, X is perceived or remembered. The converse of a relativizing predicate need not be such a predicate. The "necessary connections" Hume did not find were conceived as symmetrical. Here Hume's determinism blinded him to theoretical possibilities. Particular predecessors may be necessary to an event, even though particular successors are not.

In propositional logic we have as relations: (1) mutual independence; (2) one-way dependence—simple conditioning or entailment; (3) interdependence—biconditioning or equivalence. Only extreme monists question the genuineness of these three forms. Yet many philosophers hold that existents or actualities are subject to no similar triad of possibilities. Extreme monists deny of

¹ John Passmore, *A Hundred Years of Philosophy*, Second Edition (1966; Baltimore, 1968). See also Passmore, *Philosophical Reasoning* (1961), pp. 34f.

² See, e.g., articles "Relations, Internal and External" and "Panpsychism" in *The Encyclopedia of Philosophy*, ed. by Paul Edwards (New York, 1967). Also article "Bradley" in J. O. Urmson's *Concise Encyclopedia of Philosophy* (New York, 1960). Or E. Gellner's *Legitimation of Belief* (Cambridge, 1974), pp. 56, 66–70, 128–38. In all these cases some points I am trying to make in the present essay seem both relevant and overlooked.

³ For my philosophy the only close to definitive source is *Creative Synthesis and Philosophic Method* (LaSalle, and London, 1970). See especially chs. I, V–VI, IX, X.

actualities (1) and (2) and regard even (3) as mere appearance. Extreme pluralists accept only (1) as valid of actualities, with the trivial exception, pointed out by Russell, that a whole, i.e., a group of independent actualities, depends upon its parts or members, thus conforming to (2). One-way dependence is one-way inclusion. Yes, but the issue now becomes: Is every whole with independent constituents a mere group of the constituents? Are experiences *either* (as in New Realism) groups of mutually independent items, *or else* (as Moore held) single entities whose data are purely external to them? Are successive events simply external to one another? Russell and many other pluralists answer these questions affirmatively, thus trivializing the existential form of (2).

The *Principia* confusion (pp. 220f., 218f.) between material implication (which makes all true propositions equivalent) and entailment may have helped to mislead Russell here. Note his grudging admission (pp. 221, 219) that inferences may "practically" be "one-sided." Perhaps it is also practical to admit one-way dependencies between single actualities.

All three forms of existential dependence occur in the structure of time and space. (1) According to relativity physics, challenged by few philosophers, contemporary, spatially separated momentary events are causally independent of one another. (2) According to common sense, physics, and many philosophers, temporally separated events have causal conditions only in the past (no messages come from the future). (3) According to common sense, physics, and many philosophers, contemporary enduring things or careers, at least if nearby relative to their duration, condition each other. Thus the analogy, basic to this essay, between formal logic and ontology is significant. The disconcerting aspect of (3) (are neighbors *equivalent*?) can be overcome by relativity physics, with its analysis of things into events as final units of reality. We shall recur to this point. "Interaction" of individuals is a complex of one-way actions of unitary events. Unqualified "organic unity" is misleading shorthand, though classical physics does not suffice to make this clear.

I may seem to confuse logical and causal dependence. But my view is that where causal dependence obtains, it is, except for human ignorance, logical.⁴ In an adequate language it is analytic that events require, depend upon, their precise predecessors. Insofar I agree with the monists. How-

ever, (a) in no proper language is it analytic that events depend upon precisely specified successors. Rather, the denial of this is analytic. Here I agree with the pluralists. Also (b) logical relations between concrete events, even those to predecessors, transcend distinct human perception or insight, since such events are incomparably richer in determinations, logically "stronger" in description, than mere abstractions. No one, unless divine, could sheerly know concrete actualities. Moreover, the contingency of each step forward in time consists in this, that reality grows richer in definiteness, i.e., the full description of any situation must be logically stronger than that of its predecessors, somewhat as an adult is more complex than a child. This is the creative nature of becoming emphasized by Peirce, Bergson, and others. It is also, as I have argued elsewhere, the key to all contingency whatever, and hence (principle of contrast) to necessity as well.

Russell, like Whitehead, stresses the inadequacy of the "subject-predicate" form of proposition, in which an attribute is affirmed of a single subject, with no proper provision for relational predicates requiring two or more subjects (pp. 217, 215). Russell shows the trouble this made for Leibniz. But how does Russell treat relations? In effect he makes them entities additional to terms. These additional entities do the relating. On the contrary, I hold, as Peirce by implication does, that terms do their own relating. "There are relations" means: there are relative or dependent things (Peirce's dyads or seconds, triads or thirds, and polyads). Given an experience of *X*, it is not the relation of experiencing as such that relates the particular experience to *X*; it is simply the experience itself. "The relation" is just an abstraction from the concrete relative actualities which are experiences. If I now love *X*, it is I now, not a mere relation of loving, that enacts this relational situation. My present self relates itself to *X* and thus is conditioned by it. *X* need not be influenced thereby, and may know nothing about it. One might love someone seen only through one-way glass, or an ancestor. The conditioning may be non-mutual.

Relational words describe the relativities of existence, and these, like logical conditioning, are basically one-way. Biconditioning is a special case. Thus knowing, but not being known, conditions its subject. Plato did not know us, and that is why, whereas we know and are influenced by him, the converse influencing does not occur.

⁴ See my "Creativity and the Deductive Logic of Causality," *The Review of Metaphysics*, vol. 27 (1973), pp. 62-74.

If "substances" are to know other substances, they cannot be defined as wholly independent existents. Independence there must be—but also dependence. Moreover, in the correlation dependence-independence, apart from which neither concept has meaning, the negative sense must be derived from the positive one. In Russell's scheme dependence is taken as derivative! His single terms are simply independent, and only by considering a group of them as though they were a single term can he find the positive aspect of the correlation. Rather, the single terms must be both dependent and independent, the first with respect to one set of other terms, the second with respect to a different set.

Consider Hume's axiom, approved of by Ayer and Von Wright: "What is distinguishable is separable." I am distinguishable from my ancestors, who existed without me; do I exist without them? They "no longer" exist; still, if history is true or false of anything, events and things do not become non-events or non-things by becoming past. We require the past existence of ancestors; did they require our future existence? Some would say so, but can they give a nondogmatic reason for it? I have argued elsewhere that prophecies, unlike histories, do not get their meaning from possible correspondence with events they denote; rather their sense derives from the axiom, whatever happens tomorrow will have to show the influence of what happens today. Prophecies try to express this influence.⁵

II. RELATIONS, INTERNAL AND EXTERNAL

The debate over relations with which the century began proceeded almost entirely on the assumption that one must choose between the contrary extremes: universal internality or universal externality of relations. (In this essay, *Rab* is internal to *a* if *a* depends on the relation, hence on *b*, for its actuality, or at least for more than the mere status of being so related. The obscurity of the "at least" phrase is unavoidable if the paradigm actualities are taken to be substances rather than unit-events. This is one of the many indications that the substance concept is not the ultimate term of analysis. Rorty's view that the definition of internality depends upon how the "nature" of *a* is talked about arises from his not penetrating to the event standpoint.⁶ Of course John Jones could have been John Jones had a certain insect not flown into his field of vision at a certain moment. But the experience he

had at that moment must have been different, or rather, *that* experience that he did have would not have occurred.

Consider the following debate.

Monists: relations cannot be universally external, for then they would be additional entities, over and above those to be related, hence themselves needing to be related, and this would start a truly vicious regress; therefore relations are universally internal.

Pluralists: Relations cannot be universally internal, for then terms would be indistinguishable; therefore relations are universally external.

(The monists, at least, had additional arguments but these were not impressive).

In both cases "therefore" was unjustified. Neither avoidance of a regress nor distinguishability requires mutual, but only one-way, dependence or independence. A relation internal to one term but not the other is no additional item, but merely an aspect of the first term; yet the two terms remain distinct, since one is the dependent, the other the independent factor.

Imagine two architects.

First architect: "In my houses doors will not be hinged, for that would make doors indistinguishable from doorways; however, one edge of a door will have a relation of fixed adjacency to an edge of its doorway, and the other three edges of the door will have relations of variable adjacency to the other edges of the doorway. All these relations will be symmetrically external to their terms."

Second architect: "Your doors will fall to the floor. My doors will be hinged, but to make assurance doubly sure, they will be hinged on all four edges."

First architect: "Your doors will be immovable and useless."

Second architect: "It seems we cannot agree."

This ludicrous exchange models the way relations have been dealt with by some famous philosophers, thus on one side Ockham, Hume, Russell, Perry, Moore, Carnap; on the other side Spinoza, Bradley, Bosanquet, Joachim, Royce, Blanshard. For the four edges of doors and doorways put the dimensions of time-space; for the fixity of doorways, substitute the unalterability of past conditions for present decisions; for the partial non-fixity of doors substitute the partial openness that will be attributed to the future by those who have not fallen on their faces before old-fashioned science (summed up in the famous saying of Laplace) or old-fashioned

⁵ See my "The Meaning of 'Is going to be'," *Mind*, vol. 74 (1965), pp. 46-58.

⁶ Richard Rorty, "Relations, Internal and External," *The Encyclopedia of Philosophy*, *op. cit.*

theology (summed up in Jonathan Edwards's essay on the will) according to which every event is fully determined by previous (or eternal) conditions.

It is fair to recall that, as Bacon long ago and Einstein recently have said, reality is "subtle." Whatever view a philosopher takes, there will be apparent difficulties. Like avoiding sin (according to Reinhold Niebuhr), finding philosophic truth is "an impossible possibility." Still, three remarks are in order. (1) Errors, like Aristotelian vices, fall into contrary extremes, hence it is wise to be skeptical of extremes. (2) The truth of one extreme is not implied by the falsity of the other. From *not all* we cannot infer *none*. (3) Monistic and pluralistic arguments discredit neither external nor internal relations, but only the exclusive assertion of either in violation of the principle of contrast which (*pace* Wittgenstein) is a metaphysical rather than anti-metaphysical axiom.

What has been said above is so nearly obvious that I am not likely to be the only one who knows it. One of my teachers, C. I. Lewis, disagreed with both extremes in the relations controversy, without, as I saw it, finding the right middle ground. I can think of eight or more writers, in several countries, who have had insight into one-way relativities. One of these was W. E. Hocking, another of my teachers (Passmore's sketch of him, though perceptive, does not bring out this aspect). The clearest of the group were Peirce, especially in his category of "secondness", and Whitehead, in his category of "prehension." Yet few philosophers seem aware of the bearing of these ideas on the relations controversy. Passmore is scarcely helpful here (pp. 140, 138; pp. 344, 341).

Wittgenstein gave up the logical atomism he shared for a time (pp. 424f.) with Russell; but what view, presumably intermediate between this doctrine and an equally extreme monism, is true he naturally did not say, being the professed liberator from metaphysical doctrines. One may view standard language as implicitly wiser than Russell or Bradley without seeing in the contrast between relativizing and non-relativizing predicates a mere misuse of words. There must be dependencies and independencies, and it must be possible to talk sensibly in general terms about them.

Some years ago I asked Ewing, a mild partisan of internal vs. external relations, if he had thought of the possibility of admitting both. He said he had not. And Russell in Chicago told me that one must choose between a thoroughgoing monism and a thoroughgoing pluralism. "And," he added, "I

don't see any rational way of deciding between them." (Any reply I might have made was aborted by Russell's being called to the phone.) H. H. Price, a careful thinker, in his *Hume's Theory of the External World*, has no fault to find with Hume's "atomism". So he too is one who thinks that the monists are wrong, not because they are extremists, but because they pick the wrong extreme. Moreover, Price has religious concerns: it seems elementary (though the question cannot be discussed in this essay) that creatures are thought to be distinguishable from, but dependent upon, the Creator. This contradicts atomism.

III. REALISM AND IDEALISM

Closely connected with the relations debate, early in this century, was that concerning "realism" and "idealism." Here, too, writers seemed to take sides without adequately exploring theoretical options. Moore, in his "Refutation of Idealism" (pp. 290f., 217f.), an essay in which he later saw serious flaws (*loc. cit.*), argued, like so many others of that time, for an extreme position. In "awareness of blue," the "of" relation is taken to be external to both terms. Thus awareness is "transparent," colorless, in all cases identical, no matter how various the data. Is it not Moore's error that is transparent? What all cases of awareness have in common is indeed neutral to variations in data, but such a common denominator is a universal, not a concrete particular. What all cases of "color" have in common is similarly neutral to difference between blue and any other color.

Suppose Moore right in the other half of his dual negation: that a patch of color one is aware of is independent of that awareness, i.e., the relation does not relativize the datum. By no logical rule does it follow that it fails to relativize the awareness. Moore even regards "relative thing" as contradictory (*loc. cit.*), but he has in mind the extreme, indeed absurd, case of a thing relativized by every relation in which it stands.

In his essay on external and internal relations, Moore overgeneralizes the former. This is partly because he fails to employ the superior analytic power of a suitable event language, compared to the more usual thing language. The genetic identity of changing individuals is partly neutral (*pace* Leibniz) to internal details. I could have been "myself," Charles Hartshorne, had I never met the woman who is my wife. I should have been considerably different, however, probably much less

happy and efficient. Indeed I might have died years ago. As Buddhism has held for two millennia, and relativity physics for some decades, the most definite terms of relations are not things, persons, substances, but momentary states, unit instances of becoming, which Whitehead calls "actual entities." Russell had a somewhat similar view, and when Peirce speaks of "the logic of events" as the existential counterpart of propositional logic, or when he says that a person is an "idea," he is inclining toward without quite reaching the same view. What Russell misses is that, although the sheer "identity" of a changing thing is an absurdity, the opposite extreme of sheer diversity—mutually external successive states—is equally so. Strawson, Shoemaker, see the Russellian (Humean) failure but not the Peircean-Whiteheadian success.⁷ They fail to see that the event language is compatible with one-way dependence and hence partial identity.

S. Alexander's epistemology is like Moore's: mental act and physical object are merely "com-present" to each other. The New Realists, too (Perry *et al.*), were radical pluralists, blind to one-way relativity (pp. 262f., 260f.). Critical Realists, e.g., Santayana, had another form of the same bias. Only essences, not independent actualities, were held to be strictly given (pp. 285, 283; pp. 290, 288). Yet our naive intuition takes experiences to depend asymmetrically on, and have as data portions of, the physical world. (It will appear later that this is not quite "naive realism.") So-called realists and so-called idealists alike tended to reject this natural view in favor either of the complete disunity of subject and object or of their complete organic unity. Thus Royce, in his *The World and the Individual*, defends his universalizing of internal relations by refuting as "realism" the equally unguarded assertion of external ones (93). As extreme pluralists define realism, the refutation is cogent. But extremists make, and are partly made by, bad definitions. Royce's monistic idealism no more follows from his refutation of "realism" than "all" follows from "not none." Peirce saw this at once.⁸ He was a metaphysical idealist, but a moderately pluralistic one. Hence he could be a realist in epistemology (261, p. 260).

IV. REALISM AND PSYCHICALISM

Just as there is no forced option between all rela-

tions external and all internal, so there is none between: "we experience only private *sensa*" and "we experience entities independent not only of our experiences but of experiences of any kind, human or otherwise." A third, logically allowable possibility is that the data of any experience, though independent of that experience, are nevertheless themselves experiences of some kind. This view, often called panpsychism (pp. 287f., 285f.) is better termed *psychicalism*. The idea is not that "everything" is in the same sense psychical. As Wittgenstein remarks, a teapot is not the sort of thing that can be thought of as a subject. However, the main line of psychicalist philosophers (Fechner and Royce being illustrious exceptions) follows Leibniz in distinguishing between active singulars and the aggregates of such singulars that our nonmicroscopic vision and touch present as essentially inactive quasi-singulars. "Teapots do not feel" by no logical rule entails "molecules do not feel." Nor does "trees probably (*pace* Fechner) do not feel" justify "plant cells probably do not." Logic texts might, but do not, offer these cases to illustrate the fallacy of division. It begs the question of psychicalism to neglect the logical-type distinction between singulars and composites.

The proper argument for psychicalism has rarely been stated clearly, whether by friend or foe. It may be outlined as follows:

1. The data of human experiences, in the clearest cases, are *given* as themselves experiences, but experiences other than and independent of those to which they are given. The clearest cases are of two kinds: memories, the data being previous, especially just previous experiences of the same human person; and perceptions of bodily pains and pleasures. We shall return to the latter presently.

2. Since independence is negative, and its meaning therefore derivative, those processes which most directly illustrate the structure of reality must be the dependent ones. Realism, in the intuitively natural sense, takes experiences to be such processes. It also takes experiences to exhibit independence. For when, in immediate memory, we intuit previous experiences, what we intuit was itself one or more experiences; but these did not depend on the particular way we subsequently remembered them. Thus in memory "mind" is on both sides of the dependence-independence contrast, as it is on both

⁷ P. F. Strawson, *Individuals* (London, 1959). Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, 1968).

⁸ For Peirce's review of Royce's *The World and the Individual*, see *The Collected Papers of Charles Sanders Peirce*, ed. by Charles Hartshorne, Paul Weiss, and A. W. Burks (Cambridge, 1930-1958), paragraph 8. 126ff.

sides of the subject-object contrast. Conceived sufficiently broadly, as not restricted to the human or even animal forms, mind is, psychicalists hold (and I know of no careful refutation of the contention) capable of all conceivable varieties of dependence and independence. Even careless or strongly biased criticisms of this metaphysical view—it does not admit of empirical tests (in Popper's sense) but only of logical ones—are hard to find.

3. The only impressive advance on Hume's pluralistic analysis of causality, apart from equally one-sided monistic views (e.g., Spinoza, Blanshard), is found in a psychicalism that views instances of becoming as universally instances of experiencing whose data are previous instances. It is analytic (in a fully adequate language) that what we experience influences us. Hence, if memory is experience of the past—and what else is it?—then *either* remembering is not experienced, *or* cases of causal influence are experienced. But remembering is experienced, for we remember remembering. And so Hume is mistaken. The realistic view of memory is that it is experience of the past.

"Mistakes of memory" always involve something more, and less, than memory; they involve interpretation, verbalization, more or less unwitting guesses. To doubt that we experience the past is in principle to doubt everything. For by the time we make use of perceptions they are past events. Perception itself is reasonably viewed, e.g., by some Buddhists and Whitehead, as experience of, hence as influence by the past.⁹ So far as extra-bodily percepta are concerned this is a scientific truism. Are even inner-bodily events simultaneous with experiences of them? Not if the natural view of experiences as asymmetrically dependent upon their data is valid. Simultaneity is a symmetrical relation, and nowhere in science is a one-way influence united with a symmetrical temporal relation. Should this incongruous union be admitted in philosophy?

Alexander (pp. 269, 268) says we "contemplate" physical things and "enjoy" or live through mental states. If by contemplate he means simultaneously have as data, then nothing is contemplated. However, our mental states, once they are lived through, can be contemplated—in immediate memory. (Ryle and Whitehead notably agree in construing intro-

spection as a form of memory, of retrospection.) The options for theory are not limited to those Alexander discussed.

A telltale phrase of some realists is that objects are "outside and independent of" awareness. Often implied is the equivalence of "included in" and "dependent upon" (pp. 182, 181; pp. 212, 210, pp. 260, 258). Perry, however, shrewdly avoids this assumption. A pluralist must avoid it. Every entity is in the universe; if this makes it dependent upon the totality of entities, there is no independence. As Russell saw, it is including, not being included, that implies dependency. When given to an experience, data are in the experience, but (on the process view) they also existed previously, hence independently.

Still another failure to explore theoretical possibilities is the refusal to consider physiological conditions of perception as its most immediate data. Instead one assumed the dilemma: *either* seen or touched (mostly extra-bodily) objects or events *or* merely private *sensa* or mental states are given. That *sensa* are bodily processes has, however, been held by some (including Whitehead as well as Russell) who were well aware of objections that seem conclusive to others. More of this presently.

4. In experience are found both relational structures and qualities in the nonstructural sense, both "primary" and "secondary" properties. Science generalizes structure, but not quality, far beyond the directly given instances. Yet without quality structures are empty, incurably abstract—as Russell has admitted. Only psychicalism generalizes quality as well as structure (this is most clear in Whitehead's work) to cover nonhuman forms of reality. With Leibniz, Berkeley, and idealists or psychicalists generally, I hold that materialists and dualists will in vain seek a solution for this problem of abstractness.¹⁰

Consider physical pains and pleasures. (Examples of the latter include sexual ones, described by H. G. Wells as "like being flayed with delight.") In these cases at least it is bodily, not extrabodily processes we are most directly aware of. We feel pain when, under some other not fully understood conditions, certain cells are damaged. Pain, like physical pleasure, is as revelatory of a physical state of affairs as color sensations. Doctors, but not all philosophers, know this. However, Nunn (pp. 260,

⁹ Whitehead's alternative to Hume on causality is most accessible in his *Symbolism, Its Meaning and Effect* (New York, 1927, 1959), Ch. 2.

¹⁰ See my "Perception and the Abstract Concreteness of Science," *Journal of Phenomenology and Philosophical Research*, vol. 34 (1974), pp. 465-476.

258) argues that pains are no mere subjective states of ours. They "raise obstacles in our path", and are objects not qualities of our awareness. The positive part of this contention is to the point. On the other hand, surely pains, though objectively there, are instances of the psychical. Indeed, Ducasse is right against Moore (in the latter's Refutation): awareness of sensory qualities, whether pain, red, or bitter, is differentially characterized by these qualities.¹¹ What both disputants miss is that the subject-datum relation is one of *participation*. This is most readily seen in immediate memory. One re-feels, directly and more or less vividly participates in, his just previous feeling of anger, grief, or enjoyment. So in physical pain one intuitively grasps previous suffering which, in this case, was not one's own human feeling but the subhuman feelings that certain cells have just had.

We do not distinctly intuit cells as cells, but science has revealed a microstructure of physical reality, whether in the body or out of it, that human intuition cannot directly and distinctly grasp. (Long ago some astute Greeks guessed this must be the case. Current philosophical writers often seem pre-Democritean.) Indistinctly we must intuit physical reality; otherwise our experiences and sciences are unintelligible. Current philosophy is justified in its rejection of mere phenomenalism; the intelligible, unpopular, and not really refuted alternative is psychicalism, with the mind-body relation conceived as participatory.

The participatory aspect is the element of truth I see in Heidegger's insistence, against Husserl, that we are given to ourselves as *in the world*. Hearing and reading Husserl and Heidegger in 1924-25, and reading them since, has not convinced me that they are the best guides available, even in phenomenology, i.e., in analyzing the given as such.

The truth in Sir Karl Popper's contention (p. 412) that tables are not mere collections of atoms is, I believe, in principle contained in the admissions: (1) qualities as well as structures are given and belong to nature generally, (2) bodily cells as well as and more directly than atoms are given.

5. Although given entities cannot depend on particular experiences of them, yet in a somewhat non-Berkeleyan sense, "to be is to be perceived." Recall that the perceived precedes and conditions the perceiving, and that, according to the process

view, though conditions do not imply or depend on particular effects that result from them, yet they do imply, depend on, the truth that there must be some such effects. The world that produced Hitler might not have produced him, but it was bound to have consequences, and consequences expressive of that world. Events do not necessitate their successors, but they must have some suitable successors. This is a valid, though non-Leibnizian, principle of sufficient reason. Only psychicalism explains the "must" in this principle. "Suitable" successors means those able, at least in minimal ways, to "prehend" their predecessors.¹² Thus to be (as past) is to be perceived (also, usually, remembered). As an event is happening it is not given to any experience—though it consists of one or more instances of experiencing. But having happened it is then given to further instances. Events are contributions to their successors, experiences to subsequent experiences. This "contributionism" is a metaphysical basis for ethical responsibility. It is also in partial agreement with Wittgenstein's polemic against "private" entities or words. Human experiences begin as private, but their privacy is short-lived. Quickly they become public facts. First bodily processes, consisting of subhuman (cellular) experiences, and then less directly extra-bodily ones, inherit or prehend our experiences (in their limited subhuman ways). So far from "other mind" being unexperienced, nothing else is experienced. In the event language, even one's own past experiences are, strictly speaking, "other" for one's present experience.

V. BERGSON, PEIRCE, AND WHITEHEAD

Among those who have sensed the nonsymmetrical participatory structure of reality was Bergson, with his creative-cumulative view of becoming. Every moment, he thought, new, not fully foreseeable actualities are created. These must intrinsically, though for the most part not consciously or distinctly, refer to antecedent actualities. No one ever, perhaps, so vividly appreciated the absurdity of treating past events as nonentities, or "future events" as more than somewhat indeterminate potentialities. Yet Bergson failed to state—indeed denied the possibility of stating—the asymmetrical view in a lucidly logical way. He wrote that psychical states "interpenetrate", "melt into one another

¹¹ For Ducasse and Moore on qualities and awareness see *The Philosophy of G. E. Moore*. The Library of Living Philosophers, ed. by P. A. Schilpp (LaSalle, and London, 1942, 1952), pp. 223ff., 653-660.

¹² See note 4 above.

and form an organic whole" (pp. 106, 105). Under Lovejoy's prodding he substituted "continuity" for the other expressions, still without clarifying the asymmetry. The "prolongation" of earlier into later is better, for surely the later is not prolonged into the earlier. But nowhere does Bergson achieve clarity about the logical structure of his doctrine. As so often, the claim to dispense with logic, a claim by which he earned William James's gratitude, proved a self-fulfilling prophecy. Of course Russell was not impressed, and Peirce was led to say (pp. 105, 104), "Bergson seems to be doing his prettiest to muddle all distinctions."

A clearer headed, independent partisan of the cumulative-creative view was Peirce. He wrote: "time is . . . objective modality," "the past is the sum of accomplished facts," "the indeterminate future becomes the irrevocable past." But at certain points Peirce, too, blurs the picture. His "secondness" is defined as "being relative to . . . or in reaction with something else," in contrast to "firstness," i.e., "being or existing independently of anything else."¹³ The former definition is noncommittal between mutual and one-way dependence. The "with" suggests mutuality. In some of Peirce's examples, as in immediate memory, asymmetry is implied. In others, as when two men push a door in opposite directions, there appears to be symmetry. (Here pre-relativity physics was unhelpful.) And when in a solitary passage the question is put, which is basic, the mutual or the one-way form, no answer is given. Yet Peirce himself elsewhere points out that logical biconditioning is a complex special case of simple conditioning. What Peirce lacked, as did his contemporaries, was insight into the asymmetrical temporal structure of perception. He also, like so many before quantum physics, overestimated the continuous aspect of reality. This was the weak side of his "synechism."

Whitehead (with a different physics to help him) was, in his technical doctrines, consistently on the side of one-way dependencies as primary. "Prehension," the common factor of memory and perception, is said to be "the most concrete form of relatedness" and the prehended data to be temporally prior and independent. "There are no occasions in the future"; in the becoming of an actual entity "the many become one and are increased by one"; each entity is a "creative synthesis" of its predecessors. They are in it, it is not in them—otherwise there would be no increase in multiplicity.

(Since Bergson denies temporal multiplicity he cannot assert its increase.) Unfortunately, in some of Whitehead's nontechnical expositions, in his often eloquent rhetoric, in which he tends somewhat to overstate, he frequently writes as if he thought all relations were internal. He did not so think, but having had much to do with pluralists, he put the stress on internal relations and the "solidarity" of the world. Yet his formal explications of *prehension, contemporary, future, creativity* ("the category of the ultimate") limit the solidarity. Passmore's exposition, though perceptive on some aspects of this philosophy, is insufficiently technical to bring out its logical structure.

VI. THE PREJUDICE OF SYMMETRY

Why has it been so difficult to grasp the mixture of dependence and independence which is reality?

There is the perennial appeal of simplicity: "all relations are internal," "all are external"; each is simpler than "some are one, some the other." The classic case of this simplistic appeal is causal determinism. What precedes an event must, one thought, contain "necessary and sufficient" conditions for the event. It is more complicated to say: (1) an event's predecessors must contain conditions necessary and collectively sufficient to make the event *possible*; yet (2) the final definite actualization, out of the range of possibilities thus established, is made by the event itself (Bergson's, Berdyaev's, Whitehead's creativity, Peirce's spontaneity). On this view there are no sufficient prior conditions for actuality, but only for possibility.

"Necessary and sufficient" is symmetrical: if *p* is necessary and sufficient for *q*, then *q* is sufficient and necessary for *p*. Yet our natural intuition is that results depend on antecedent factors, not vice versa. Classical causality made becoming like a logic with only equivalent propositions. Physics, preceded by Philosophy—Bergson, Meyerson (pp. 331, 329), Peirce, Dewey, and others—has been moving away from this excess of symmetry; but many philosophers seem still enamored of the Newtonian idea. Otherwise, there would have been less effort to show its compatibility with freedom.

The simplicity of symmetry is deceptive. *A* and *B* may "love each other," but *B*'s love for *A* is no mere duplicate of *A*'s love for *B*. "Cause and effect are equal"—but perhaps they never quite are. The inequalities may for some purposes be neglected—

¹³ Peirce, *op. cit.*, par. 6.32; also 1.24f.; 5.66, p. 469.

which explains the power of classical physics—but for all we could ever know empirically they are always there. Perhaps only in the negative case, mutual non-dependence, is there literal symmetry. (If Bell's theorem in quantum physics is correct, even this negative symmetrical case does not occur.) But this applies to spatially, not temporally separated items. Space is the order of two-way, time of one-way, dependence and independence. Strictly predictable effects are abstract outlines (approximations or statistical frequencies), as Maxwell, Boutroux, and Peirce guessed in the last century. Science must be abstract, but it seems to move closer to the real asymmetries—e.g., in ideas of entropy, cosmic expansion, chance variations (more literally so in neo-Darwinism), "propensities" (Popper), half-life laws, non-parity, the origins of life (Prigogine), the incomplete symmetry between electric and magnetic charges (Schwinger). I have heard Hoyle (despite his quasi-symmetrical steady-state view) and the physicist Rabi agree that the future, but not the past, is in principle and apart from human limitations incompletely knowable (as determinate).

Another form of the lure of simplicity is revealed in the query, must not complex entities have wholly simple constituents? (Recall Locke's simple ideas.) If so, there is no one-way relativity among these constituents, for dependent things must be complicated by their relata; and so, even if the latter are entirely simple, the former cannot be.

As further source for symmetrical pluralism, there is the trust in introspection. Thus Russell and Lovejoy cannot find the whole past in present awareness, and no one else (unless God) can consciously detect it there, item by item. But then experiencing, and knowing that and what one experiences, are far from identical. With Leibniz, Peirce, Bergson, Freud, and psychologists rather generally, I hold that our introspective powers are very limited. But they do suffice to show that our awareness is more or less incurably indistinct.

There is the argument, used, e.g., by Russell: past events are not now occurring, hence they cannot be constituents of present process. It is assumed that there is but one way for an event to be real now. And this is the question at issue. The view of process philosophy is that the actuality now becoming is a nascent subject whose immanent objects are past events which, as they became, were incomplete

subjects not yet objects for any further subject. Being past, being object, is a relation, and this relation, as qualifying particular present subjects, is external to the entity which is past. Actualities "become, they do not change"—but are preserved in subsequent actualities. For us now past events are dim, shadowy, scarcely real objectivities, but this dimness qualifies us, not the past events, nor yet all present actuality. Human awareness is not so simply "the measure of things." (For theists, including "process" theists, deity is the measure.)¹⁴

There is the idea, common to Hume, Russell (pp. 238, 236), and other proponents of the "identity theory" of mind and neural states, that the subject-object duality can be collapsed into a single nonrelative (i.e., "absolute") type of entity, called by some "neutral" (pp. 263, 261; pp. 262, 263), by Hume "impression," by James (pp. 108f., 106f.) in his confused "radical empiricism," "pure experience," by still others "brain states." Richard Rorty (personal communication) credits Wittgenstein (as Russell did James) with helping us to rid philosophy of the subject-object relation. The clearest counter example is memory, of which Hume and his followers give no credible analysis. The data of sense experience are, I think, neural, but having the data is not. Data do not in this sense "have," receive, intuit themselves. Nor do I accept Rorty's interpretation of Wittgenstein—or Russell's (or even, in some of his moods, James's) of James.

How odd that Russell, who in spite of his one-sided metaphysics said so many penetrating things, among them, "asymmetrical relations are the most relational of relations," yet missed one-way dependency—the decisive relation, since it is the basis of inference. Peirce's theory of categories turns on it; Popper makes it the one primitive of his formal logic (408, p. 406). It was not Russell who developed the philosophical bearings of the "Logic of Relatives" that he took over from Peirce (pp. 139f., 137f.; pp. 221, 219). It was first Peirce himself and then Whitehead who at least began this development. (To find Whitehead clearer than Peirce and Peirce than still others on certain issues is not to prejudice topics outside the scope of this essay. Process philosophy is uncommitted to either man, whether against the other or in broader comparisons).

To the asymmetrical creative view of becoming it might be objected that it seems to commit us to an

¹⁴ For the Process View of deity see Whitehead's *Process and Reality* (New York, 1929, 1957), Part V, and *Creative Synthesis*, *op. cit.*, chs. XI–XIV.

infinite regress of previous events with still earlier events as their data, and thus to one horn of Kant's antinomy about past time. I do not see how to avoid this implication. There is one comfort: Moore, with a very different philosophy, defends the idea of an actually infinite past.¹⁵

Bosanquet remarked ironically on the "meeting" of philosophical extremes, but he missed part of the irony, since only by loose thinking could he view his monism as less extreme than the doctrines he opposed (p. 85f.). (Panpsychism no more "turns all things into persons" (p. 88) than locating persons

in space turns all things into rocks. Of all modern views, psychicalism grants the most variety to nature: besides human and animal feelings, it admits protozoon, molecular, atomic, subatomic ones; also insentient things, i.e., loosely organized macro-objects, such as rocks and even trees.) Still, doctrinal opposites do, as Bosanquet says, coincide in some defects, including his and the pluralists' joint blindness to one-way, creative, and participatory dependencies. So long as this blindness persists, evaluation of those authors called philosophers of process will be less than adequate.¹⁶

The University of Texas

Received February 5, 1976

¹⁵ G. E. Moore, "The Notion of Infinity," in *Some Main Problems of Philosophy* (London, 1953; New York, 1962, 1966).

¹⁶ Bell's philosophically exciting theorem, even when freed of its needless assumption of hidden variables, implies that if certain very subtle influences, incapable of conveying messages, are taken into account, spatially separated events are not wholly independent of one another, and that not all influences are limited to the speed of light. Although no physical means are known to decide between the subtle influence in question going from *a* to *b*, or from *b* to *a*, or both ways, the doctrine held by many physicists and philosophers that causes precede effects rules out the two-way option.

As the physicist Stapp has seen, process philosophy can now be revised and strengthened by the elimination of symmetrical independence, the idea of relations that are nowhere—a paradox that has troubled me for several decades—without falling into the contrary paradox of mutually internal relations between unit events, a view that makes existents indistinguishable from one another and violates the true part of Hume's axiom, what is distinguishable is separable or independent (nonmutually). Interaction between persons or things (not unit events) can still be accepted, as commonsense requires.

See H. P. Stapp, "Bell's Theorem and World Process *Nuovo Cimento*, Vol. 29 B, N. 2 (11 Oct., 1975), pp. 270-276. Also Stapp, "Theory of Reality," Lawrence Berkeley Laboratory, University of California, Berkeley, 94720 (FBL 3837).

VI. PRISONERS, PARADOX, AND RATIONALITY

LAWRENCE H. DAVIS

I

A TYPICAL statement of what is called the "prisoner's dilemma" runs as follows:

Two men suspected of committing a crime together are arrested and placed in separate cells by the police. Each suspect may either confess or remain silent, and each one knows the possible consequences of his action. These are: (1) If one suspect confesses and his partner does not, the one who confessed turns state's evidence and goes free and the other one goes to jail for twenty years. (2) If both suspects confess, they both go to jail for five years. (3) If both suspects remain silent, they both go to jail for a year for carrying concealed weapons—a lesser charge. We will suppose that there is no "honor among thieves" and each suspect's sole concern is his own self-interest. Under these conditions, what should the criminals do?¹

This case and others like it contain an element of paradox. There is a compelling *basic argument for confessing*, even though *if neither confessed, the outcome would be better for each*. In this version, if both confess they each go to prison for five years, but if they are both silent, they each go to prison for only one year. Rationality, represented by the basic argument for confessing, leads them to an outcome less attractive in the self-interested view of each than another outcome which lies in their grasp, so to speak. Apparently in these situations, "rational pursuit of self-interest generates an outcome detrimental to sheer self-interest," as Max Black has recently put it.² Closer examination shows however that the matter is quite complex. In particular, it seems there is no genuine paradox, and the basic argument for confessing may be inapplicable to a theoretically important class of cases. An implication of this would be that co-operation between individuals with clashing interests may be more rationally defensible

than has been widely thought; but this implication will not be explored here.

II

To begin with the basic argument for confessing, in a sentence it is that each prisoner is better off confessing than maintaining silence, no matter what the other does. Let us represent the four pairs of alternatives and the associated outcomes in the standard way, abstracting from details of the example which shall not concern us.

		Column	
		do not do <i>A</i>	do <i>A</i>
Row	do not do <i>A</i>	(<i>b</i> , <i>b</i>)	(<i>d</i> , <i>a</i>)
	do <i>A</i>	(<i>a</i> , <i>d</i>)	(<i>c</i> , <i>c</i>)

Here the two individuals are named "Row" and "Column," and the choice each faces is whether to do *A* (in the example, "maintain silence") or not to do *A* (in the example, "confess"). The outcome for each pair of alternatives is represented in the matrix by an ordered pair whose members indicate how the four outcomes are ranked in the judgment of Row and Column, respectively. The outcome (*d*, *a*), for example is rated "best" by Row and "worst" by Column; the outcome (*c*, *c*) is rated "second best" by each.

Now consider from Row's point of view what happens if Column does not do *A*. If Row also refrains from doing *A*, the outcome is (*b*, *b*), Row's third best. But if Row does *A*, the outcome is (*a*, *d*), his worst. If Column does not do *A*, then, Row does better if he follows suit.

If Column does *A*, then if Row does not do *A* the outcome is the one he rates best, and if Row does *A*, the outcome is only his second best. Again, he does better if he does not do *A*. In the standard phrase, not doing *A* *dominates* doing *A* from Row's point of view. The conclusion seems inescapable that Row should not do *A* (in the example: he should confess).

¹ Morton Davis, *Game Theory: A Nontechnical Introduction* (New York, 1970), p. 93.

² Max Black, "The 'Prisoner's Dilemma' and the Limits of Rationality," forthcoming. I am indebted to this paper and numerous discussions with Professor Black for the inspiration of this study and many details along the way; but he should not be held responsible for my conclusions.

Since the positions of Row and Column are exactly symmetrical, parallel reasoning yields the conclusion that not doing *A* is the rationally prescribed alternative for Column also. If both are rational enough to see and accept this argument, the outcome is the one each ranks "third best," even though if each does *A*, the outcome is the one each ranks "second best."

III

Why, exactly, is this result supposed to be paradoxical? R. D. Luce and H. Raiffa find it "slightly uncomfortable that two so-called irrational players [i.e., individuals in this sort of situation] will both fare much better than two so-called rational ones,"³ and others have shared this discomfort. But rationality has never been supposed to be a *guarantee* of the best possible outcome; at least it would be a mistake to suppose that it is. An agent forced to choose between an hour alone with a rattlesnake or a lion, who may be hungry or sleeping, may be best advised to choose the snake. If the snake bites him, he can still get medical attention later; but the lion can kill him easily within an hour. For all that, it may be that the lion is asleep on a full stomach and the agent suffers snakebite "needlessly."

If there is paradox here, it arises from the fact that each agent faces neither a lion nor a snake, but another rational agent like himself. And although they rank the outcomes differently, they do agree in rating the outcome of their both doing *A* above the outcome of their both not doing *A*, to which their rationality apparently condemns them.

Even so there would be only tragedy but not paradox if neither agent *knew* whether he faced a rational agent or a dangerous animal, or simply an irrational agent. And now we notice a curious omission in some presentations of the prisoner's dilemma and the theory of games in which it is discussed. The version quoted specifies that "each one knows the possible consequences of his action," but only that "we" suppose there is no "honor among thieves." Now, it is natural enough to suppose that each has this information about the other's motivation. Or as Luce and Raiffa (pp. 54f; cf. pp. 47-50) make conveniently explicit for game theory in general,

Each player is assumed to know the rules of the game in full detail [and] also the payoff functions of the other players.

Each player is assumed to be "rational" in the sense that, given two alternatives, he will always choose the one he prefers, i.e., the one with the larger utility.

This still does not quite say that each knows the others have this knowledge; nor does it say that each knows the others are "rational" in the indicated sense. Taking this silence as an indication of the contrary, it appears that in Luce and Raiffa's view, we do *not* assume the players have this knowledge about one another, in which case the "discomfort" they feel about the prisoner's dilemma does not arise from any genuine paradox.⁴

Other writers feel differently. J. C. Harsanyi speaks of "mutually expected rationality" and the "mutual perception of each other's rationality"; Anatol Rapoport refers to Harsanyi, and goes on to assert that "the fundamental assumption of game theory is that everything there is to know about a situation is known at the start by 'rational players' . . ." Davis (pp. 101f) also cites Rapoport's view that arguments like the basic argument for confessing given in section II depend on the assumption "that one's opponent will act rationally," and does not disagree. For writers making assumptions like these, there may indeed be something paradoxical about the basic argument for confessing. In fact, explicitly making a strong assumption about the players' knowledge of one another's rationality and knowledge enables an apparently valid argument for *not* confessing! To this argument we now turn.

IV

The basic argument for confessing was that "each prisoner is better off confessing than maintaining silence, no matter what the other does." Whatever Column does, the outcome is better as Row ranks outcomes if he does not do *A* than it is if he does *A*; not doing *A* is a dominant strategy for Row. Now phrases such as "*whatever* Column does . . ." suggest that Row is uncertain what Column will do, and it may be that applicability of the argument—relevance of dominance considera-

³ R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, 1957), p. 96.

⁴ On p. 5 they speak of the need to make explicit all "assumptions about [each player's] ability to perceive the game situation."

⁵ J. C. Harsanyi, "On the Rationality Postulates Underlying the Theory of Co-operative Games," *Journal of Conflict Resolution*, vol. 5 (1961), pp. 179-196, and "Rationality Postulates for Bargaining Solutions in Cooperative and Non Cooperative Games," *Management Science*, vol. 9 (1962), pp. 141-153; Anatol Rapoport, *Two-Person Game Theory* (Ann Arbor, 1966), pp. 139-141; cf. also pp. 54f. 60, 70, 103f.

tions—sometimes depends crucially on there being this uncertainty.⁶

If we assume that *each prisoner knows that each knows that each is rational*, as well as knowing the information represented in the matrix in section II, it appears that Row is *not* uncertain what Column will do. More precisely, he can and will determine what Column and he himself *should*, and so by assumption *will*, do. He knows that Column is a rational agent and that he himself is. He knows further that their situations and information are symmetric: whatever considerations would lead Row himself to choose one way would lead Column to choose *exactly the same way*. The information Row has, then, implies that *the outcomes (d, a) and (a, d) are not in fact possible*. Of the two outcomes remaining, (c, c) is rated higher by both Row and Column. The rationally prescribed alternative for each, then, is to do A—maintain silence. And by assumption, each will take this alternative, secure in the knowledge that the other will take it also.

This may be called the *basic argument for co-operating*; a somewhat more formal version is presented in section VI, below. It will emerge that the only part of this argument open to serious question is the emphasized clause to the effect that two of the outcomes are not “possible”; in section VII I try to clarify and defend the relevant sense of “possibility.” Other questions about the argument are dealt with in the next section.

Before starting on these questions, however, it is worth noting that the principle of maximizing expected utility may seem to yield the same conclusion where the assumption is made that each prisoner knows that each knows that each is “rational” in this strong sense, which entails that each will take the alternative rationally prescribed for him. For we have from Row’s point of view:

$$\text{E.U. (confessing)} = [u((b, b)) \times \text{pr}((b, b)/\text{Row confesses})] + [u((d, a)) \times \text{pr}((d, a)/\text{Row confesses})].$$

The utilities of (b, b) and (d, a) from Row’s point of view are *b* and *d*, respectively. It may be thought that $\text{pr}((b, b)/\text{Row confesses})$ is equal to one, since

by assumption Row and Column each take the rationally prescribed alternative: if Row confesses, then confessing must be rationally prescribed, in which case Column also confesses, and the outcome is (b, b). If this is correct, then we also have $\text{pr}((d, a)/\text{Row confesses}) = 0$, and the expected utility of confessing is precisely *b*. Similarly, the expected utility of maintaining silence is *c*, and so Row will maintain silence.

A problem with this is that there is another, perhaps equally plausible, way of interpreting the probabilities here. Perhaps “ $\text{pr}((b, b)/\text{Row confesses})$ ” should be read as the probability that the outcome would be (b, b) if Row *were* to confess, where it is allowed that the antecedent of the conditional may be contrary to fact. If a “fact” to which “Row confesses” may be contrary is the assumed fact that he takes the rationally prescribed alternative, we cannot reason as above that the probability is one. There may be other problems with using the principle of expected utility in this way, so we shall not consider it further.⁷

V

1. In his *Two-Person Game Theory*, Rapoport presents an argument the final portion of which is virtually identical with the basic argument for co-operating. It will pay to quote it in full, adjusting the notation to coincide with ours:

The best outcome for both of us is (c, c). However, if Column assumes that I shall do A, he may well not do A to win the largest payoff. To protect myself I will also refrain from doing A. But this makes for a loss for both of us. Two rational players certainly deserve the outcome (c, c). I am rational and by the fundamental postulate of game theory, I must assume that Column is also rational. If I have come to the conclusion that doing A is the rational choice, he too must have come to the same conclusion. Now knowing that he will do A, what shall I do? Shall I not refrain from doing A to get the greatest payoff? But if I have come to this conclusion, he has also probably done so. Again we end up with (b, b). To insure that he does not come to the

⁶ More than this certainly needs to be said about the force of dominance considerations and their limitations, but I will not try to do so here. One reason is that a full treatment would require examination of such other topics as Newcomb’s paradox. (See Robert Nozick, “Newcomb’s Problem and Two Principles of Choice” in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht, 1969), pp. 114–146.) I will only assert dogmatically my belief that the “dominance principle” which bids one to employ a dominant strategy if available is *not* a fundamental principle of choice such as the principle of maximizing expected utility may be. Its applicability apparently depends on assumptions about the choice situation normally taken for granted in the usual matrix representations of these situations, assumptions which Newcomb’s paradox and the class of cases discussed here call into question. Perhaps this discussion will make this claim somewhat plausible.

⁷ The contrary-to-fact conditional re-enters the discussion briefly in a different context, at the end of section VI.

conclusion that he should refrain from doing *A*, I better avoid it also. For if I avoid it and am rational, he too will avoid it if he is rational. On the other hand, if rationality prescribes not doing *A*, then it must also prescribe not doing *A* for him. At any rate because of the symmetry of the game, rationality must prescribe *the same choice to both*. But if both choose the same, then *(c, c)* is clearly the better. Therefore I should choose to do *A*.⁸

Prior to the sentence beginning "On the other hand . . .," Row argues back and forth, from the wisdom of doing *A* to the need to protect himself against Column's *not* doing *A* to the realization of the folly of their *both* not doing *A* and so on—a process that threatens to be endless. The chain of reasoning in the final four sentences is not a continuation of what precedes, but an entirely fresh approach, which leads to a definite conclusion. But the key assumption, that each knows (or assumes) that each is rational, and has precisely the same relevant information as the other, plays an explicit role in both chains of reasoning.

J.W.N. Watkins, acknowledging Rapoport, has made use of an argument essentially the same as that of the *first* part of Rapoport's, and his use of the "key assumption" is criticized by A. K. Sen:

Furthermore, the type of egoistic reasoning outlined by Watkins essentially makes each prisoner assume—at least temporarily—that the other prisoner's action will be a *function* of his own action and will in fact coincide with it. . . . It is precisely because [Column's] choice cannot be assumed by [Row] to be a mirror-reflection of his own choice that the dilemma of the prisoners is supposed to arise.⁹

But Sen does not here defend the alleged impossibility of Row's assuming Column's choice to be a "mirror-reflection" of his own. To the contrary, it seems an obvious entailment of the assumption that each knows that each knows that each is rational, together with the symmetry of the situation. And I have argued in Section III that the prisoner's dilemma is paradoxical only if we make this assumption.

The assumption licenses each to assume the other will act precisely as he himself will act only if "being rational" is understood in a very strong way. Occasionally Luce and Raiffa seem to imply that a "rational" agent *does not make mistakes*, and although they should not be held to this implication, anything weaker would dissolve all element of paradox.¹⁰ Assume for the moment that the basic argument for cooperating is valid. If Row thinks of the argument and has no worries about Column's "rationality," then it is indeed rational for him to go ahead and do *A*. But if he has the slightest doubt—even if his doubt is that Column might *fail to think* of the argument, or might *make some mistake* in assessing it—then it is by no means so obvious what he should do. Well, *if* he is really limited to the information represented in the matrix, then the basic argument for confessing (not doing *A*) presumably applies. If he knows what the outcomes are, as well as their rank ordering, then depending on the details it may be "rational" for him to gamble. Either way, his situation is unfortunate and perhaps difficult—and in this sense a "dilemma"—but there is no *paradox*. Our interest then is only in cases where each knows (or is willing to act on the assumption) that each is "rational" in the *quite* strong sense of not making mistakes *and* not failing to think of relevant arguments. This may remove the discussion from reality, but should not remove its theoretical interest.

2. The basic argument for cooperating may seem to *abuse* the assumption about their rationality and knowledge, by having each agent think his [free?] choice *determines* what is rational, and thereby determines what the other agent chooses. This impression is fostered by Rapoport's having his agent say "two rational players *deserve* the outcome *(c, c)*" and "to *insure* that he does not come to the conclusion . . . I better avoid it also." Watkins also has his agents say:

But dammit, we *don't* have to resign ourselves to *(b, b)*. We have been offered the chance of *(c, c)*. Let us grab it.¹¹

⁸ Rapoport, *op. cit.*, pp. 141f. I learned of Rapoport's argument, and the variant presented by Watkins (see next note), several days after developing the argument of section IV.

⁹ Amartya K. Sen, "Reply to Comments" in Stephan Körner, *Practical Reason* (New Haven, 1974), p. 80. Sen is here responding to Watkins' "Comment: 'Self-interest and Morality,'" which is itself a set of comments on Sen's "Choice, Orderings and Morality." All three are in the same volume, pp. 54–82. Watkins' original presentation of his argument is in his "Imperfect Rationality" in Robert Borger and Frank Cioffi (eds.), *Explanation in the Behavioural Sciences* (Cambridge, 1970), pp. 202–206.

¹⁰ In the passage quoted above, they say a rational agent "will always choose the [alternative] with the larger utility," but earlier they say merely that a player "will attempt to maximize expected utility" (emphasis added. See also pp. 50 and 80.)

¹¹ Watkins, in Körner, *op. cit.*, p. 73. In his earlier presentation in the Borger and Cioffi volume, however, Watkins merely has his agent say "I should choose [A], confident that he, by similar reasoning, will choose [A]." (Borger and Cioffi, *op. cit.*, p. 205.)

and this suggestion of a *causal* or other *determining* relation between the two choices may be part of what Sen finds objectionable in each agent's assuming that the other's action "will be a *function* of his own." But the choices and actions of each may be a function of what is rationally prescribed for each, rather than of each other. (Cf. two clocks known to work perfectly and to have been synchronized at an earlier time. By looking at one we can know what time the other indicates, which is not to say we can *make* the other one indicate an arbitrarily selected time by setting the first to that time.) The misleading passages do not occur in the second part of Rapoport's argument or in the argument of section IV, above, and nothing like them is presupposed.

3. If valid, the basic argument for co-operating overcomes the following apparent circularity. "To conclude that the outcome of doing *A* is better than the outcome of not doing *A*, Row must know what these outcomes are. This depends on what Column does, which by assumption depends on what is rationally prescribed for him (and Row also). But *this* depends on which action has the better outcome. Row must know the answer in order to arrive at the answer!" The way in which the argument purports to overcome this circularity may be clarified by listing its premisses:

- (1) An alternative *X* is rationally prescribed for an agent *y* if *y* knows that there are just two possible outcomes *m* and *n*, such that if *y* takes *X* then the outcome is *m*, if *y* does not take *X* then the outcome is *n*, and *m* is better (in *y*'s judgment) than *n*.
- (2) Each prisoner knows that each knows that each will take the rationally prescribed alternative.

From the background information supplied by the matrix and the fact that (2) assigns the same information and rationality to each, it can be inferred that an alternative is rationally prescribed for one just in case it is for the other. From (2), it can then be inferred that each *does* just what the other does. What the two possible outcomes are can now be determined first on the assumption that doing *A* is rationally prescribed and then on the

assumption that not doing *A* is rationally prescribed, and so it can be determined that *A* satisfies the sufficient condition given in (1) for being rationally prescribed, for each. (See section VI, below.)

The "trick," then, is to reason about "the rationally prescribed alternative" before it is known which it is. But this presupposes *that* there is one. Worse, (2) presupposes further that each agent *knows* there is one. What may be called the "cyclic aspect" of the reasoning in the first part of Rapoport's argument, repeated in Watkins' version, suggests the conclusion Watkins in fact draws, that the agents "are in a dilemma; there is no determinate solution to their optimization-problem."¹² If this is correct, (2) cannot be true. And even if it is not correct, (2) can be true only if the agents can *know* there is a unique rationally prescribed alternative before they reach the conclusion of an argument based on (2).

But it is easy to show there exists a unique rationally prescribed alternative. Assume there is none. Then all our strong assumptions about the rationality of the agents and their knowledge of their rationality leaves each still unable to determine what the other will do. (This is the situation portrayed in the first part of Rapoport's argument.) But if Row, say, cannot determine what Column will do, then the basic argument for confessing applies, and the rationally prescribed alternative for him is to confess. (In this situation, "confessing" satisfies a sufficient condition for being rationally prescribed other than the one given in (1)).

But then there is a rationally prescribed alternative for him, contradicting the assumption that there is none. Row and Column can know, then, that there is a rationally prescribed alternative and be confident they will each take it before they have figured out what it is; (2) can be true, at least so far as this issue is concerned.

4. The question remains, what is wrong with the "cyclic aspect" of the reasoning in the first part of Rapoport's argument (and in Watkins')? For unless there is some error, we may expect that the second part of Rapoport's argument, our basic argument for co-operating, is itself but the first link in another endless chain, in which case (as just argued) the rationally prescribed alternative is after all to confess.

¹² Watkins, in Körner, *op. cit.*, p. 73. Rapoport also seems unwilling to say definitely that the rationally prescribed alternative for each is cooperation. Shortly after the passage quoted he writes "Either the concept of rationality is not well defined in the context of the nonnegotiable nonzero-sum game; or if the definition of rationality in the context of the zero-sum game is applied to the 'solution' of some nonzero-sum games, the results are paradoxical." His conclusion seems to be that the concept of rationality is not well defined in these contexts. See his remarks on pp. 142ff.

One subsidiary point is that the word "cyclic" is really inappropriate, since the reasoning would *not* continue indefinitely as both Rapoport and Watkins seem to think. The "cyclic aspect" would at best simply be a *reductio* of the assumption or any attempted argument that the rationally prescribed alternative is to maintain silence. (See steps (9) to (12) of the more formal presentation in section VI, below.)

The main point however is that what is rationally prescribed for an agent is *relative to the relevant information he has*, and this means *all* the relevant information. If all that Row knows is that Column will do *A* (maintain silence), then it is certainly rational for him *not* to do *A*, yielding the outcome (*d, a*) which he rates as optimal. But Row is never in the situation of knowing or thinking *merely* that Column will do *A*. "Column will do *A*" occurs in his reasoning only as an inference from premises including or implying the premise that it is rational for *both* of them to do *A*. Thus we quoted Rapoport's agent as saying:

If I have come to the conclusion that doing *A* is the rational choice, he too must have come to the same conclusion. Now knowing that he will do *A*, what shall I do?

But Row "knows" this only if he also "knows" that doing *A* is the rational choice, and in fact the one that he will make. His question "What shall I do?" has already been answered, understood as "What is rationally prescribed for me relative to all the relevant information I have?" Understood as "What is rationally prescribed for me relative *simply* to my knowledge that he will do *A*?" the question is irrelevant.

VI

This discussion of the "cyclic aspect" may be better appreciated if we attempt a more formal statement of the basic argument for co-operating and the attempted *reductio* of its conclusion. We begin with the premises listed above, but will not attempt to include all the steps and background assumptions.

- (1) An alternative *X* is rationally prescribed for an agent *y* if *y* knows that there are just two possible outcomes *m* and *n*, such that if *y* takes *X* then the outcome is *m*, if *y* does not take *X* then the outcome is *n*, and *m* is better (in *y*'s judgment) than *n*.
- (2) Each prisoner knows that each knows that

each will take the rationally prescribed alternative (recall the independent argument that there is just one rationally prescribed alternative).

- (3) Each knows that an alternative is rationally prescribed for one of them just in case it is also rationally prescribed for the other of them.
- (4) Each knows that he will maintain silence just in case the other does, and that he will confess just in case the other does.
- (5) Each knows that if silence is rationally prescribed and he maintains silence, then the outcome will be (*c, c*); and that if confessing is rationally prescribed and he confesses, then the outcome will be (*b, b*).
- (6) Each knows that (*c, c*) and (*b, b*) are the only possible outcomes.
- (7) Each knows that he judges (*c, c*) to be better than (*b, b*).
- (8) Therefore, maintaining silence is rationally prescribed for each.

Rapoport and Watkins would presumably have the argument continue as follows:

- (9) Each prisoner knows that maintaining silence is rationally prescribed for each.
- (10) Each knows that the other will maintain silence.
- (11) Each knows that if he were to confess, the outcome would be one he judges better than (*c, c*).
- (12) Therefore, confessing is rationally prescribed for each.

This result contradicts (8), showing there must be a flaw somewhere in the argument. But it is not clear that the flaw occurs prior to (8); to the contrary, the step from (11) to (12) is invalid, if it is supposed to be based on (1). For (1) would be applicable only if the outcome judged better than (*c, c*)—either (*d, a*) or (*a, d*), depending on the agent—were known to be "possible," along with (*c, c*). But this is not known, and is in fact contradicted by (6), according to which (*c, c*) and (*b, b*) are the *only* possible outcomes. The argument leading to (6) makes use of (2), whereas a claim that (*d, a*) or (*a, d*) is "possible" seems to require averting one's gaze from (2). The prisoners themselves could think that (*a, d*) or (*d, a*) was "possible" only if they ignored the information which (2) asserts they possess; but surely it is out of place for agents to ignore relevant information in the course of deliberating. This is the point made at the end of the preceding section.

One response is to deny that the step to (12) is based on (1). It may be claimed that:

- (1') An alternative X is rationally prescribed for y if y knows that the outcome that would result if he took X is one he judges better than the outcome that would result if he did not take that alternative

is at least as plausible as (1), and warrants the move from (11) to (12).¹³ But if (1) is satisfactory as it stands, with its notion of "possible outcomes," then it seems clearly superior to (1'). For in terms of this notion of "possible outcomes," (1') would have to be understood in something like the following sense:

- (1'') X is rationally prescribed for y if y knows that there are outcomes m and n such that if y were to take X , then m would be possible and in fact would be the outcome; if y were not to take X , then n would be possible and in fact would be the outcome; and m is better (in y 's judgment) than n .

Thus (11) should be understood as saying that each prisoner knows that if he *were* to confess—contrary to the facts as described in (10)—then one of the outcomes ruled out by (6) *would be* both possible and actual. But why should consideration of what "*would be possible if*" be relevant to determining what is rationally prescribed for an agent in the actual world?

The claim must be then that (1') is superior to (1) because the latter is wholly *unsatisfactory*, employing a notion of "possible outcomes" which *cannot* be clarified. This, rather than any of the other objections considered in section V, is the challenge facing the basic argument for co-operating.

VII

At first blush, the relevant notion of "possible outcomes" may seem perfectly clear. Each prisoner is assumed to know that each will take the rationally prescribed alternative, and to know that their situations are symmetrical in all relevant respects. "Obviously" it cannot happen that they do different things, so only two outcomes are "possible": that which would result from their both confessing, and that which would result from their both maintaining silence. Perhaps nothing more *need* be said

to defend the notion and hence the basic argument for co-operating; the problem is the difficulty of finding much more which *can* be said.

Perhaps every coherently describable outcome is "possible" in some sense; e.g., it may be *logically* possible that the prison walls dissolve before the prisoners announce their decisions. But this is not the desired sense. Assuming each prisoner has the *ability* to confess, and the ability to remain silent, there is a sense in which each outcome represented in the matrix is "possible." For we naturally move from " y has the ability to take alternative X " to " y 's taking X is possible," and so from " y has the ability to take X and z has the ability to take W " to " y 's taking X and z 's taking W are (jointly) possible." But this also is not the desired sense. In the desired sense, there are two and only two "possible" outcomes, not as many as there are entries in the matrix, and not as many as there are coherently describable ones.

The desired sense is *epistemic*, and *relative*: the outcomes (b, b) and (c, c) are said to be alone possible *relative to each agent's relevant information*. The word "relevant" is necessary here, for example in case they already know how they will act. (A reliable oracle told them, or they have gone through this argument before; I assume this sort of knowledge is possible and would not block deliberation.) If they know how they will act, they know what the outcome *will be*, and relative to *this* bit of information, no *other* outcome can be considered "possible." Exactly what is to be included in and excluded from the body of "relevant" information may be a troublesome problem in its own right, but we shall assume it solved and focus on another.

The desired sense of "possible" is *epistemic*. Roughly, if on the basis of his relevant information, a rational agent will realize that a certain one of the outcomes represented in the matrix will not be the actual outcome, then that outcome is *impossible* relative to that information. Let ' p ' stand for a sentence conjoining all the relevant information of one of the agents. If we assume that an ideally rational agent knows everything entailed by everything he knows, we might try defining "possible outcome relative to p " by

- (D) (x) (x is a possible outcome relative to p if and only if

- (i) x is an outcome; and

¹³ This is essentially a suggestion made by Robert Stalnaker. I am indebted to him for several helpful discussions and illuminating criticisms of the argument of this paper; again, responsibility for my conclusions is solely my own.

- (ii) p does not entail that the outcome is not x .)

We can understand (i) as " x is represented in the matrix." Ordinarily, an agent's information is not enough to rule out any of these outcomes, so all remain "possible." But as noted in the preceding paragraph, if the agent knows how each will act, then only one outcome is possible relative to this information:

Unfortunately, if the argument from (1) to (8) (supplemented by the unstated background assumptions) is deductively valid, then if (D) is accepted it again follows that only one outcome is possible relative to p . For (8) and (2) entail that each will remain silent, and together with (5) this yields the conclusion that the outcome will be (c, c) . This means that the argument is *not* deductively valid, since it requires (at step (6), needed for the applicability of (1)) that (b, b) also be "possible." To avoid this result, (D) must be abandoned or altered.

Now actual agents do *not* know everything entailed by everything they know, and it is a mistake to attribute this knowledge to the "ideally rational" agents of our example and of game theory in general. (Alternatively, certain of the things entailed could be excluded from the agent's "relevant" information.) Clause (ii) of (D), then, is too strong. Perhaps an ideally rational agent could *come* to know any given thing entailed by some proposition he knows, if he reasoned "in the right direction," so to speak. This suggests replacing (ii) with

- (ii') an ideally rational agent could *not* infer from p that the outcome is *not* x .

This will not do, for essentially the same reason as before. If the basic argument for confessing is valid, then an ideally rational agent *could* infer (8), hence that the outcome is not (b, b) , from p .

What we must note is that a reasoning process takes *time* and involves *steps*. We have no comprehensive and convincing theory of the nature and individuation of the objects of thought and belief. Let us imagine that an agent reasons from one *sentence* to another. In our example, we suppose the agent starts with p (which incorporates (1) and (2)) and ends up with (8); along the way, he reasons to (6). And the point is that he reaches (6) *before* he reaches (8) or an equivalent in his reasoning. This suggests

- (ii'') an ideally rational agent could *not* infer

from p , before inferring for some w that w is the rationally prescribed alternative for him, that the outcome is *not* x .

as a substitute for (ii) in (D). After he reaches (8), the agent may realize that (b, b) also cannot be the outcome. That is, there may perfectly well be a sense in which his relevant information entails that there is only one "possible outcome," namely the one which will be actual. But since he cannot rule out (b, b) until after he concludes that maintaining silence is rationally prescribed, there is a sense in which he can correctly label it "possible" while engaged in the reasoning which finally leads him to that conclusion. This is the sense in which (6) declares that (b, b) as well as (c, c) is a "possible outcome."

But is it really *true* that an agent "cannot" rule out (b, b) until after he reaches (8) or an equivalent? If it is possible at all for him to validly infer (8) from the relevant information he starts with, hence also to validly infer that (b, b) will not be the outcome, it may be that the latter can be validly inferred, reached by a chain of valid reasoning, without using (8) or any obvious equivalent as a "steppingstone." I have not devised a precise substitute for (D) which guards against this possibility, and it may be doubted that one can be devised that would do this. The problem of clarifying the notion of "possible outcome" employed in (6) remains unsolved, and the basic argument for co-operating remains neither refuted nor conclusively established.

VIII

In this section I summarize the main conclusions of the preceding sections, and suggest a few applications and directions for further inquiry.

1. The "Prisoner's Dilemma" may be a dilemma, but presents no strict *paradox* unless (2) is assumed; and if the basic argument for co-operating is valid, there is no paradox even if (2) is assumed.

2. Some standard discussions of the "Prisoner's Dilemma" have left it unclear whether or not (2) was to be assumed, but some would apparently insist on (2). Yet (2) is implausible in the extreme for any real-life analogues of the Prisoner's Dilemma. To know or assume that an agent is "rational" in a sense that he *will* without fail take the alternative rationally prescribed for him is to know or assume that he already knows which alternative that is, or can be relied upon to figure it out in the finite time available for deliberating, even if there is no known

algorithm for doing so. There must also be no "slip-up" between his awareness of the rationally prescribed alternative, and his taking it—no "weakness of the will" or sudden physiological malfunction. But (2) requires that both agents know or assume that they are both "rational" in this super-strong sense.

3. Despite the implausibility of (2), there is some point to considering its implications in Prisoner's Dilemma and other contexts. It is needed for speculation à la Hobbes or Rawls as to what ideally rational agents would do in a "state of nature" or other "original position." Hobbes in particular would seem to be vindicated if the basic argument for co-operating is valid.¹⁴ Similarly, the argument bears on discussions of the rationality of voting: if valid, it shows that it is rational for ideally rational citizens to vote if they believe that enough others are equally rational.¹⁵ True, none of us are in such a situation, and so *in practice* it may often be irrational to vote (ignoring secondary reasons agents

usually have). But mere knowledge of the theoretical result should soothe the misgivings of some over the rationality of participation in democratic institutions. More generally, the argument, if valid, shows that co-operation is rational from a self-interested point of view in some contexts in which it was thought not to be. As both co-operativeness and egoistic rationality have been touted as virtues, it cannot help but be beneficial to know that they are not so irreconcilable.

4. But is the basic argument for co-operating valid? So far as the matter has been pursued here, the answer depends on a notion of "possible outcome" which resists formal definition. This may show the argument invalid. Alternatively, it may show that there is something about the idealized notion of a rational agent which just cannot be captured by a formal theory. The latter would have significant implications for the whole enterprise of formal decision theory, and it is to be hoped that the correct verdict is neither of these two extremes.¹⁶

University of Missouri—St. Louis

Received May 19, 1976

¹⁴ Watkins, in "Imperfect Rationality," argues that rational agents in a Hobbesian state of nature could never get out of it, in accordance with his view that there is no rationally prescribed alternative for agents in these situations.

¹⁵ More exactly, the others must not only be "equally rational," but also convinced of one another's rationality and convictions. The argument is then that all will behave alike; the only "possible outcomes" are that in which all vote who can, and that in which none vote. Of these two, the first is presumably judged better by each, so is rationally prescribed for each.

¹⁶ This paper was written while I was a fellow in the Humanities, Science, and Technology Unit of the Cornell University Program on Science, Technology, and Society.

VII. IDENTITY CONDITIONS FOR EVENTS

MYLES BRAND

WHAT is the nature of events? Some philosophers have recently argued that they are propositions (or states of affairs or facts), in that they are abstract and repeatable.¹ Others have argued that events are particulars, and are relevantly similar to physical objects.² This latter view has *prima facie* advantages. It reflects our linguistic practices of talking about events and physical objects in the same ways; for example, we use quantification, definite and indefinite descriptions, count-expressions and proper names similarly in both cases. More importantly, a particularist view of events is ontologically parsimonious, in that the theory requires the least kinds of existents. Proponents of the view that events are propositions need abstract entities, and the property of occurring or obtaining, neither of which is required by a particularist theory.

The key to advocating a particularist account of events—or any account of events—is to provide adequate identity conditions. I argue for this methodological point in Section 1. In the remainder of the paper, I present and defend the view that, with important qualifications, necessary spatio-temporal coincidence provides adequate identity conditions for events. Physical objects and events are alike in that they are, essentially, spatio-temporally locatable.

1. IDENTITY CONDITIONS

Confusion in the recent literature has developed because the function of identity conditions has not been delineated explicitly. Actually, under the rubric "identity conditions," statements intended for three different functions have been proposed.

Sometimes one of these statements is intended as a criterion for judging or telling how many of a certain kind of object there are, or for re-identifying a single object of a kind. When the function of the statements are epistemological, it seems best to label them "*identifying conditions*." Consider the claim that *S* and *T* are the same person just in case they have the same parents and siblings. Clearly, this is an inadequate statement of identity conditions: since parents and siblings are also persons, it is circular. Nevertheless, this statement is sometimes useful for making practical judgments about whether there is one or more persons. Or consider the view that *S* and *T* are the same person just in case their bodies occupy the same spatio-temporal regions. Although more plausible as a proposal for identity conditions than the previous statement, it is at best controversial: arguments are often adduced to show that, for one reason or another, two persons can occupy a single body. Those arguing this line are pointing to the need to specify coincidence of mental states, such as memories, beliefs and goals. If this rejoinder is correct, coincidence of spatio-temporal regions does not provide conditions logically sufficient for personal identity; but, nevertheless, it might provide identifying conditions for persons. A plausible criterion for judging or telling whether there is one or more persons, or whether it is now the same person as before, is bodily coincidence. I take it that this epistemic criterion is used, for example, in legal situations.

Quine has emphasized that the machinery of quantificational logic can be utilized only if the members of the domain can be individuated. It is sometimes thought that identity conditions are needed to satisfy Quine's requirements. However,

¹ See Roderick Chisholm, "Events and Propositions," *Nous*, vol. 4 (1970), pp. 15–24; "States of Affairs Again," *Nous*, vol. 5 (1971), pp. 179–190; "Problems of Identity" in M. Munitz (ed.), *Identity and Individuation* (New York, 1971), pp. 3–80, also see Neil L. Wilson, "Facts, Events and Their Identity Conditions," *Philosophical Studies*, vol. 25 (1974), pp. 303–321.

² The main proponent of the particularistic theory is Donald Davidson; see especially his "The Logical Form of Action Sentences" in N. Rescher (ed.), *The Logic of Decision and Action* (Pittsburgh, 1967), pp. 81–95; "The Individuation of Events" in N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel* (Dordrecht, 1970), pp. 216–234; "Events as Particulars," *Nous*, vol. 4 (1970), pp. 25–32. Also see my "Particulars, Events, and Actions" in M. Brand and D. Walton (eds.), *Action Theory* (Dordrecht, forthcoming). Although it is controversial, Jaegwon Kim's view can be classified as particularistic: see his "Events and Their Descriptions" in N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*, op. cit., pp. 198–215 and "Events as Property Exemplifications" in Brand and Walton, *Action Theory*, op. cit. Alvin Goldman's account in *A Theory of Human Action* (Engelwood Cliffs, 1970), pp. 1–19, is similar to Kim's: cf. my review in *The Journal of Philosophy*, vol. 69 (1972), esp. pp. 250–254.

in order to use quantificational machinery, it is only necessary to have the most general individuating conditions that apply to the type of objects in the domain. Suppose that the domain consists of persons. Leibniz's Law, that individuals x and y are the same just in case they have all the same properties, suffices for the purposes of quantifying over this domain. Indeed, Leibniz's Law suffices whenever the domain consists of particulars, no matter what kind of particular is involved. But the suggestion that the criterion for personal identity is to be found in Leibniz's Law misses the mark. The task of providing identity conditions for persons consists in naming properties that alone are sufficient for personal identity; more specifically, the task is to pick out key—or essential—properties of persons. Leibniz's Law *does* individuate persons, but does not do so by indicating those properties that are uniquely sufficient for the identity of persons. Similarly, Leibniz's Law individuates events, but is insufficiently informative. The concern in the debate about identity conditions for events, as in the debate about personal identity, involves more than satisfying the prerequisites for using quantificational logic; it involves specifying the nature of events, specifying those characteristics unique to events.

The function of a universal identity statement central to debate about events is the ontological one of specifying the nature of a kind of object. Quine is most likely concerned with this function when he says "no entity without identity," though he seems to have confounded the function of supplying prerequisites for quantificational logic with this one. For example, in considering sets, he is not satisfied with Leibniz's Law, but rather seeks conditions peculiar to sets, namely, sameness of membership.³ In the remainder of the discussion, I shall reserve the label "identity conditions" for universal statements of identity purporting to specify the nature of a kind of object.

But what is it for identity conditions to specify the nature of a kind of object? My conjecture is that identity conditions do this by specifying non-trivial essential properties of objects of that kind.

Consider sets, a case in which the identity conditions are clear: $A=B$ iff $(\forall x)(x \in A \equiv x \in B)$. In fact, this statement is not quite correct, since A and B are presumed to be sets. The correct statement is

- (1) $(\forall A)(\forall B)$ [if A and B are sets, then $A=B$ iff $(\forall x)(x \in A \equiv x \in B)$].

It is essential to sets that sameness of membership is sufficient for identity. Hence, the following stronger claim also obtains:

- (2) $(\forall A)(\forall B)$ {if A and B are sets, then $\Box [A=B$ iff $(\forall x)(x \in A \equiv x \in B)$].

Abstracting the form of this statement yields

- (P) $(\exists F)(\forall x)(\forall y)$ [if x and y are ϕ 's, then $\Box (x=y$ iff Fxy)].

In the case of sets, the relational property F is having the same members as. For the sake of illustration, suppose that personal identity consisted only in sameness of memories and beliefs. Then, for ϕ =person, the property F would be having the same memories and beliefs as. Note that (P) does not require that the identity conditions be as inclusive as Leibniz's Law, nor that they function epistemologically as identifying conditions.

It is important to be clear about the essential property indicated by statements of the form (P). It is a property of pairs of objects which belong to the sort ϕ . It is not a property essential to each object of the sort, not one such that an object belongs to a sort in virtue of having it. This latter type of property, expressed by a monadic predicate, is related to the dyadic property indicated in (P), but not as a matter of logic alone. In the illustration about personal identity, there is a relationship between " S has the same memories and beliefs as T " and " S has memories and beliefs"; the latter monadic property, if true of persons, is not so simply because the former, relational property—having the same memories and beliefs as—holds between some pairs of persons. The case of sets clearly shows that these two properties are not related by logic alone: " A has members" does not follow from " A has the same members as B ," since it might be that $A=B=\Lambda$. In general, the monadic property semantically corresponding to the dyadic property in (P) is interesting, in that it seems to directly tell us the nature of instances of ϕ ; unfortunately, true identity conditions do not guarantee that instances of ϕ have this monadic property, nor that, if they have this property, they have it essentially. Identity conditions specify the nature of sorts of objects, but in the roundabout way of specifying a relational property which is sufficient for identity of pairs of these objects.

There is no mechanical procedure for finding the appropriate dyadic property F , given a sort ϕ , which results in a correct, nontrivial instance of

³ W. V. O. Quine, *Set Theory and Its Logic* (Cambridge, 1963), pp. 28 ff.

(P); that is a substantive issue to be decided for each sort ϕ . Some conditions of adequacy for F can be stated, however.

One obvious adequacy condition is:

(A1) F is nontautological,

where a tautological dyadic property is one that every pair of objects share. Another adequacy condition, perhaps also obvious, is that

(A2) F is philosophically neutral with respect to ϕ .

Identity conditions specify properties essential to objects of a kind; these conditions focus on the internal structure of the kind, as it were. If a philosophical thesis makes reference to something other than the kind in question, then the truth of the purported identity conditions, stated in terms of the property F , should not depend on the truth of that philosophical thesis. However, I know of no simple and precise way to explicate (A2). The difficulty concerns the notion of a philosophical thesis: what some would take to be a philosophical thesis, others would take to be an obvious truth (or falsehood).

A more formal adequacy condition is that, for a given ϕ , F is the widest property that would make a statement of the form (P) true. Consider the illustration of personal identity again. Suppose that F were not having the same memories and beliefs as, but rather the complex property of having the same memories, beliefs, and spatio-temporal locations as. The resulting instance of (P) is true; but this statement does not specify an essential property that states the nature of persons. Given the supposition of the example, that sameness of memory and belief is sufficient for personal identity, this narrower property misleadingly suggests that spatio-temporal coincidence is required for personal identity. Let us indicate this adequacy condition by saying that

(A3) F is maximal with respect to ϕ .

Given ϕ , this condition obtains if: (i) substitution of F into (P) yields a true sentence; and (ii) there is no other property F^* such that substitution into (P) yields a true sentence and, necessarily, F^* is wider than F . (A property, say G , is wider than another property, say H , just in case, for any x and

y , if x bears H to y , then x bears G to y , but not conversely.) This adequacy condition is more clearly—though less precisely—stated directly in terms of identity conditions: identity conditions I for ϕ 's are to maximal, that is, for every I^* , if I^* are identity conditions for ϕ 's and I entails I^* , then I is I^* .

More than one property can be maximal with respect to a given ϕ . Intuitively, however, only one dyadic property can provide correct identity conditions for a sort ϕ . Arguments concerning whether spatio-temporal coincidence or whether sameness of mental states, such as memories, provide correct identity conditions for persons are not vacuous; they exemplify, in part, the underlying rationale that not both can be correct statements of identity conditions for persons, that there can be only one correct statement of identity conditions. In order to capture this intuition, let us say that

(A4) F is unique with respect to ϕ .

Adequacy condition (A4) is to be taken to mean that, if there are properties $F_1, F_2, \dots, F_n$ such that each satisfies (P) for a given ϕ and every F_i is independent of each F_j (i.e., every F_i is neither wider nor narrower than each F_j), then at most one of $F_1, F_2, \dots, F_n$ provides correct identity conditions for ϕ .

2. DAVIDSON'S IDENTITY CONDITIONS FOR EVENTS

Davidson has stated identity conditions for events in terms of sameness of cause and effect.⁴ That is, letting $e, f, \dots$ range over events, " $\dots C$ —" be the relation " $\dots$ causes—", and assuming universal closure,

(I1) $e=f$ iff $(\forall g)(gCe \equiv gCf) \ \& \ (\forall h)(eCh \equiv fCh)$.

This statement has a high degree of initial plausibility: sameness of cause and effect appears to be a property essential to the sort events.

As stated, (I1) leaves open the question of what may be substituted for the variables $e, f, \dots$. Without this additional information, ($\exists$) functions as a schema for identity conditions. One way to proceed is to follow Kim in taking the canonical form of an event description to be " x 's F ing during (or at) t ."⁵ Only complex singular terms in canonical form are to be substituted for the event variables. Natural language event naming descriptions

⁴ "The Individuation of Events," p. 231. (Also see *Synthese*, vol. 22 [1970], p. 502).

⁵ See for example Kim, "Events as Property Exemplifications." For simplicity, only events involving a single object are considered. As Kim points out, this account of the canonical form can be generalized to events involving more than one object and n -adic predicative expressions. Note that adopting this canonical way of describing events does not commit one to Kim's proposal for identity conditions: see Section 5 below.

often contain definite descriptions; let us, then, interpret this canonical form flexibly and permit definite descriptions to replace 'x' or 't'. So, "history's worst tyrant raising his arm at the beginning of the battle," as well as "Hitler's raising his arm at noon," are permissible substitutions. Judging from Davidson's many examples, he too countenances complex singular terms containing definite descriptions of the object or time as event descriptions. On this way of proceeding, "my favorite event" and "the second event of the hour" are not permissible substitutions. Purely referring singular terms, such as these or, even, "George," can be allowed as substitutions, but the resulting account would be more complex without being more enlightening.

Davidson says "this criterion may seem to have an air of circularity about it, but if there is circularity it certainly is not formal."⁶ A number of persons, however, have argued that (I1) is viciously circular.⁷ The definiens requires quantification over events, but quantification over events in the definiens is permissible only if there are identity conditions independent of (I1). But Davidson is correct: there is no vicious circularity. This objection results from confounding the requirements for quantification with those of identity conditions. Since events are construed as particulars, Leibniz's Law suffices for the purposes of using quantificational machinery; it is not necessary to have *identity conditions* independent of (I1) to quantify over events in the definiens.

There is, however, a serious difficulty for (I1). Let an ineffectual event be one that has no causes or effects. According to (I1), all ineffectual events are identical, since the definiens is trivially satisfied. But that is counterintuitive, since there might be more than one ineffectual event. Consider a related criticism. Suppose that there is a causal chain in which an object first undergoes fission and then is reunited by a process of fusion. Assume further that no other object causally interacts with it before this process begins until it concludes. Now, two events occur from the time slightly prior to the fission to the time slightly later than the

fusion, where each event involves a distinct spatio-temporal object. Nevertheless, these two events have exactly the same causal histories and futures. But (I1) has the counterintuitive consequence that these events are identical. Philosophers concerned with personal identity have recognized and discussed these situations; and in any case, such situations are logically possible.

Davidson has defended himself against these objections by claiming that universal causal determinism is necessarily true,⁸ which has the consequence that there cannot be ineffectual events nor, presumably, events not undergoing causal interactions for a period of time. However, it is highly unlikely that universal causal determinism is necessarily the case. More importantly, Davidson's reply is counter to the restriction that identity conditions should not depend for their truth on some philosophical thesis. Proposal (I1) purports to specify an essential property of the sort events; but it is a property that depends on something other than events, namely, on necessary causal lawfulness. At best, (I1) specifies an accidental relational property of events, and hence, does not specify a property of events that indicates their nature.

3. IDENTITY CONDITIONS FOR EVENTS

Identity conditions for physical objects can be given in terms of sameness of spatio-temporal regions—though this claim is not wholly unproblematic. It might be suggested, then, that identity conditions for events be in terms of sameness of spatio-temporal regions, since events too are particulars.⁹ That is, letting r range over spatio-temporal regions, ". . . W —" be ". . . occurs within—", and assuming universal closure,

$$(I2) \ e = f \text{ iff } (r) (eWr \equiv fWr).$$

Note that (I2) does not require events to occur within continuous spatio-temporal regions. The playing of a World Series, for example, consists in a number of spatio-temporally separated constituent events (namely, playing of the individual games).

⁶ Davidson, *op. cit.*

⁷ For example, Neil L. Wilson, "Facts, Events and Their Identity Conditions," *op. cit.*, pp. 303–305. This objection was also made by Fred Feldman in his oral presentation of "Kripke on the Identity Theory" at the Eastern Division APA meetings, December 1974; the original paper is printed in *The Journal of Philosophy*, vol. 71 (1974), pp. 665–676, and by myself in "Particulars, Events, and Actions," *op. cit.*

⁸ In correspondence.

⁹ A suggestion of this kind was made by E. J. Lemmen in his "Comments on D. Davidson's 'The Logical Form of Action Sentences'" in N. Rescher (ed.), *The Logic of Decision and Action*, *op. cit.*, pp. 98–100.

However, (I₂) is inadequate. Two or more events can occur in exactly the same spatio-temporal regions. For example,

- (3) My swimming the channel (now)
- (4) My catching a cold (now)

are distinct, despite occurring at the same time and place.¹⁰ Similarly, "it's raining in Chicago (now)" and "it's growing dark in Chicago (now)" describe, in the context of an appropriate background story, two events occurring in a single spatio-temporal region. One more such case is:

- (5) This metal sphere's rotating (now)
- (6) This metal sphere's becoming warm (now).

Though not himself advocating (I₂), Davidson suggests a defense by claiming that (5) and (6) are not distinct, since each is identical with motions of molecules.¹¹ But this defense is not satisfactory: (5) and (6) might each specify something identical with molecular motion, but it does not follow that they are each identical with the same molecular motion. The molecular motion that appears macroscopically as the sphere's rotating is different from the molecular motion that appears macroscopically as the sphere's growing warmer.

The revision of (I₂) required to account for multiple events in the same spatio-temporal region and to differentiate events from physical objects is (with a modification to be added):

$$(I_3) e=f \text{ iff } \square (\forall r)(eWr \equiv fWr).$$

According to (I₃), (3) ≠ (4), since I could have caught the cold someplace else and could have swam the channel while remaining healthy. Similarly, (5) ≠ (6), since the sphere could have rotated in the same spatio-temporal region without growing warmer. The identity conditions stated by (I₃) entail those stated by (I₂), but not conversely.

The identity conditions (I₃) for events exemplify the same underlying rationale as those for physical objects, namely, that particulars are spatio-temporal objects. But it is important to emphasize that (I₃) does not provide adequate identity conditions for physical objects. As indicated earlier, adequate identity conditions must be maximal. If the variables *e* and *f* in (I₃) were taken to range over physical objects, the resulting statement would not

be maximal with respect to physical objects, since there is another statement of identity conditions (namely, (I₂), altered with variables ranging over physical objects) which is entailed by the altered (I₃).

Identity conditions (I₃), however, are not wholly unproblematical. It might be objected that there is difficulty in assigning spatio-temporal regions to events. Strictly speaking, this is not a difficulty for (I₃); (I₃) states identity conditions, and does not purport to be, in addition, identifying conditions. It need not be the case that judgments about the locations of events be able to be made in every instance in order for (I₃) to be a correct statement of identity conditions.

Be that as it may, all events take time—at least in the ordinary sense of "event" in question here. If an event involves physical objects, then it occupies at least all the space occupied by the physical objects during the time of the event. If the event does not involve a physical object, such as a flashing of lightning or a magnetic field increasing in strength, the event occupies at least all the space in which the change takes place. But difficulty arises for complex spatially or temporally discontinuous events, for example, my phoning my mother long-distance.¹² Where does this event take place? At the locations at which my mother and I are located? Are the locations of the telephone wires, the switching devices, the operator to be included? In general, problems of locating events are weakened by taking ". . . W—" to be ". . . occurs within —", rather than some relation requiring precise location, such as ". . . exactly occupies—". Cases involving events having eventual parts, such as my phoning my mother, are similar to cases of locating complex physical objects. Where, for example, is my house located? Does it include the empty space within the house? The detached garage? The space between the house and the garage? It would seem that a solution to the problem of locating complex events parallels that of locating complex physical objects, since both are particulars.

A consequence of (I₃) is that there are more events than Davidson recognizes. Consider

- (7) Jones' raising his arm at noon =
Jones' signaling to the chairman at noon.

¹⁰ This case is similar to Davidson's in his "Reply to Comments," in N. Rescher (ed.), *The Logic of Decision and Action*, *op. cit.*, p. 117.

¹¹ "The Individuation of Events," *op. cit.*, pp. 230–231.

¹² These cases were brought to my attention by Alvin Goldman.

Assume that Jones in fact accompanied his signaling to the chairman by raising his arm. It does not follow, however, that his raising his arm and his signaling are one and the same event. Since signaling by raising one's arm is a matter of convention, there are, to use the common idiom, possible worlds in which Jones raised his arm but did not signal and there are possible worlds in which he signaled without raising his arm. Hence, by (I₃), (7) is false. There are rather, two events that occurred in a single spatio-temporal region. Although "Jones' raising his arm" and "Jones' signaling to the chairman" are distinct, they are not unrelated. Jones signaled *by* raising his arm, or in Alvin Goldman's terminology, Jones' raising his arm conventionally generated his signaling.¹³

4. A MODIFICATION

Compare the following pair of event descriptions

- (8) Millard Fillmore's dying
- (9) the 13th President's dying

with this pair:

- (10) Hesperus' rising today
- (11) Phosphorus' rising today

Now (10)=(11) since, in every possible world, each occurs in the same spatio-temporal region. In possible worlds in which Hesperus does not exist, Phosphorus does not exist, and hence, they trivially occupy the same spatio-temporal regions in those worlds. Kripke labels proper names such as "Hesperus" and "Phosphorus" (and some other designators) *rigid*, meaning that reference is frozen, that is, they name the same object in every possible world in which that object exists and name no objects in other possible worlds.¹⁴ According to my intuitions—though I imagine not everyone's, (8) and (9) name the same event.¹⁵ However, there is a possible world where (9) did not occur but in which (10) did. For "the 13th President" is not a rigid designator and hence picks out individuals other than Fillmore in some worlds. That is, counter-intuitively, (I₃) yields (8) ≠ (9).

Before amending (I₃) so that it satisfies these intuitions, let me say and emphasize that those who find informal possible world semantics unilluminating, or who have some other favorite semantics, should feel free here and elsewhere to make the appropriate translations. The substantive point, that adequate identity conditions for events must be modal, is independent of the informal semantical interpretation given to modal notions.

David Kaplan has suggested an operator, *Dthat*, that transforms descriptions into terms rigidly designating the object picked out in the actual world.¹⁶ In brief, *Dthat* transforms descriptions into proper names; it *freezes* the reference of descriptions. This device can be adopted for our purposes. Recall that the canonical form of an event description is "*x*'s *F*ing during (or at) *t*." By applying the *Dthat* operator to the *contained* descriptions, we obtain a complex description of an actual event that is frozen across possible worlds with respect to object and time. Substituting the *Dthat*ed description into (I₃) yields the wanted intuitive results.

Stated generally, let α be a complex singular term containing one or more descriptions or names occurring transparently, and let α^+ be the result of applying *Dthat* to the names and descriptions contained in α . Then, if all the descriptions and names contained in α are rigid, α is α^+ ; if some or all the descriptions and names contained in α are not rigid, then α^+ is exactly like α except where α contains nonrigid descriptions, α^+ contains rigid designators picking out the objects designated in the actual world by the names or descriptions contained in α .

Identity conditions (I₃) are, then, to be amended as follows, where the event variables take as instances complex singular terms in canonical form:

$$(I_3^+) \quad e = f \text{ iff } e^+ = f^+ \\ e^+ = f^+ \text{ iff } \Box (r)(e^+ Wrf + Wrf)$$

Thus, for example, (I₃⁺) yields the intuitive result that (8) = (9), for (8) = (8⁺) and

$$(9) = (\text{the 13th President's dying}^+),$$

¹³ *A Theory of Human Action*, *op. cit.*, pp. 20 ff.

¹⁴ Saul Kripke, "Naming and Necessity" in Davidson and Harmon (eds.), *Semantics for Natural Languages* (Dordrecht, 1972), pp. 253-355.

¹⁵ Goldman, for example, seems to share these intuitions when he says "... two expressions may refer to the same act-token even though they contain different, nonsynonymous phrases for the same agent. If 'John' and 'the mayor' designate the same agent, then 'John's killing Smith (at *t*)' and 'The mayor's killing Smith (at *t*)' designate the same act-token" (*Theory of Human Action*, *op. cit.*, p. 12).

¹⁶ Cf. Kaplan's unpublished mimeo "Dthat." My use of this operator might be at variance with Kaplan's recent views, but it is consistent with Kripke's account (*op. cit.*, p. 346, n. 22).

and, by (I₃⁺),

(8⁺) = (the 13th President's dying⁺).

Note that the predicative components of a complex singular term are unaffected by the *Dthat*ing operation; as specified here, *Dthat*ing affects only the contained names and descriptions.

A consequence of (I₃⁺) is that, once *Dthat*ing the complex event descriptions, all true event-identity sentences are necessary. Informally, *Dthat*ing contained names and descriptions in the terms of event-identity sentences makes these contained terms rigid designators. Contained verb phrases are invariant across possible worlds, given that properties are understood nonextensionally (for example, as functions from possible worlds to sets of individuals). Hence, if an event-identity sentence whose terms have been *Dthat*ed is true in the actual world, it is true in every possible world, that is, it is necessary. For example, "Fillmore's dying = (the 13th President's dying⁺)" is true and non-contingent. It is *not* a consequence of (I₃⁺) that every true event-identity sentence is necessary. "Fillmore's dying = the 13th President's dying" is true but contingent, since "the 13th President" names some individual other than Fillmore in some possible worlds. This sentence would be necessary only if the right-hand term were *Dthat*ed.

5. COMPARISON WITH KIM'S PROPERTY-EXEMPLIFICATION VIEW

According to Jaegwon Kim, an event is an object's having a property during a time (restricting attention to "monadic events"), which is expressed in his notation as "[x, F, t]." Identity conditions are formulated in terms of identities of the constituents: assuming closure,

(I₄) [x, F, t] = [y, G, t'] iff x = y & F = G & t = t'.

That is, e = f just in case e and f involve the same object exemplifying the same property during (or at) the same time. Hence, it is essential to events that sameness of constituents is sufficient for identity.

There are *prima facie* problems for Kim's view. Among them, this view must be supplemented by a criterion for distinguishing between properties that are properly constitutive of events and those that

are not. Actually, the problem is twofold. First, properties that cannot be constitutive of events, such as being self-identical or being equal to $\sqrt{2}$ must be separated from those that might be constitutive of events. Second, properties constitutive of events must be distinguished from properties of events. For example, an event can have as a constituent the property of running; but it is not clear on Kim's account whether an event can have as constituents the properties of running quickly, running with a sword in your hand, or almost running. Kim's view must also be supplemented by a criterion for property identity; otherwise, the identity conditions for events are merely pushed back a step. Another, less noticed problem is that some events do not involve physical objects. Leaving aside the controversial case of mental events, there are changing weather conditions, changing light conditions, changing fields, and so on.

The last problem can be solved by taking spatio-temporal regions, rather than physical objects, as constituents. This solution involves problems of locating events that are similar to those confronting the necessary spatio-temporal coincidence view. With regard to property identity, it is probably too much to ask that Kim provide adequate identity conditions for properties in order to maintain the view that events are property-exemplifications; it suffices to say that solutions to a number of central philosophical problems—for instance, the mind-body problem, scientific theory reduction and meaning change—also require identity conditions for properties. The initial problem mentioned, specifying properties that can be constitutive of events, seems to me the most recalcitrant. Arguing, for example, that properties whose exemplifications are changes is not correct, since these properties do not exhaust those that might be constitutive of events. Some events are unchanges, as it were: for instance, my refraining from standing and remaining seated during the playing of the national anthem is an action in which no overt change takes place.¹⁷

In any case, let us compare the property-exemplification view, expressed by (I₄), with the necessary spatio-temporal view, expressed by (I₃⁺). Kim's view, it might be suggested, is committed to the existence of properties, whereas the particularist view is not. However, this difference may only

¹⁷ Kim has recently made some comments in "Events as Property Exemplifications" about the second part of the problem, predicate modification, but he has said little about the first part, that of isolating properties that are at least candidates for constituents of events.

be apparent. On the one hand, Kim's view about events is compatible with a reductionistic theory of properties, for example, a theory that construes properties as collections of particulars. On the other hand, a particularist theory is probably committed to properties, since, as I have argued elsewhere, an account of event recurrence requires them.¹⁸

But there is one ontological difference. On Kim's view, events are dependent on physical objects. Since events *are* physical objects having properties during times, there could be events only if there are physical objects. On the necessary spatio-temporal coincident particularist theory, events are not ontologically dependent on physical objects; there are, rather, two kinds of particulars having ontological parity.

One interesting question is whether Kim's (I₄) is coextensional with (I₃⁺). For the sake of comparison, let us agree with Kim about the internal structure of an event. On (I₃⁺), the terms naming the subject and the time are treated as rigid. Hence, the two statements of identity conditions are coextensive if and only if property identity is coextensive with necessary spatio-temporal coincidence of property exemplifications. Since, for any properties *F* and *G*, *F*=*G* only if, necessarily, all their instances are coincidental, the only case in which counterexamples can arise are those in which instances of *F* and *G* are spatio-temporally coincidental but *F* and *G* are distinct. That is, (I₄) and (I₃⁺) are coextensional only if the following principle is true, where *x* ranges over physical objects, *r* over spatio-temporal regions:

(R) If $\Box (\forall x)(\forall r)(x \text{ has } F \text{ at } r \equiv x \text{ has } G \text{ at } r)$, then *F* = *G*.

This principle might be challenged. First, the antecedent of (R) can be trivially satisfied when *F* and *G* cannot have instances, even though *F* and *G* are distinct. For example, (R) is falsified when *F* is squaring the circle and *G* is proving that a Euclidian triangle has less than 180°. Another type of case is one in which *F* is uttering all and only valid first-order schemata and *G* is uttering all and only schemata correctly derivable in a complete system. A third case is one in which *F* is, say, drawing an equilateral triangle and *G* is drawing an equiangular triangle. All the places at which equilateral triangles are drawn are places at which equiangular ones are drawn, and conversely. But

drawing an equiangular triangle is distinct from drawing an equilateral one, since some events, say using a protractor, are causally relevant to one event—drawing an equiangular triangle—that are not causally relevant to instances of the other.

However, I have some hesitancy about the falsity of (R). The first case, in which *F* and *G* cannot have instances, might be dealt with by restricting (R) to contingent properties, that is, properties such that there can be objects that have them and such that there can be objects that lack them. The second case might be dealt with by further restricting (R) to properties that can be exemplified in this world. No actual person, I take it, can utter all and only valid first-order schemata or utter all and only schemata correctly derivable in a complete system. These restrictions on principle (R) can also be applied to Kim's statement of identity conditions, since properties constitutive of events are contingent properties that can be exemplified in the actual world. The third case is more troublesome. It might be held that the property of drawing an equilateral triangle is identical with the property of drawing an equiangular one, and that if using a protractor is causally relevant to instances of one, it is also causally relevant to instances of the other. But this retort, no doubt appealing to some, begs the question. Unless there is good reason for holding that drawing an equilateral triangle and drawing an equiangular one are not distinct, the claim that instances of each property enter into different causal chains is not defeated. Since, however, there is also room for the charge of question begging from the other side, it might be best to say that this case is unresolved.

In short, if it is assumed, first, that the internal structure of events (in the monadic case) is an object's having a property during a time and if, second, principle (R), when restricted, is true, then Kim's identity conditions (I₄) are coextensional with (I₃⁺), the necessary spatio-temporal coincidence conditions. Both of the assumptions are, however, controversial. But even if both assumptions are accepted, and thus (I₃⁺) and (I₄) are taken to be coextensional, progress has been made. One problem facing Kim's view is providing identity conditions for properties. If (I₃⁺) and (I₄) are coextensional, that problem has been bypassed, since (I₃⁺) does not require supplementation by conditions for property identity. Another problem for Kim's view is that (I₄) does not give us sufficient information

¹⁸ "Particulars, Events, and Actions," *op. cit.*

to determine the nature of events; or stated another way, (I₄), to be informative, must be conjoined with a criterion for specifying properties properly constitutive of events. Conditions (I₃⁺) need not be conjoined with such a criterion, since it says that the nature of events consists in their being identical in virtue of necessary spatio-temporal coincidence. Progress has also been made in stating the identity conditions as (I₃⁺), rather than the

supposed coextensional (I₄), in that the relationship between identity conditions for physical objects and events becomes obvious. Methodologically, this is helpful because it suggests parallel treatment of problems about events and physical objects, problems such as locating events and recurrence. This last point is important, and should lead to advances in theorizing about events if properly exploited.¹⁹

University of Illinois at Chicago Circle

Received September 7, 1976

¹⁹ I have profited greatly from discussions with Daniel Berger on many matters in this paper, especially those of the last section. I also have had the benefit of valuable comments by Dorothy Grover, Richard M. Martin, Paul Teller, Irving Thalberg Jr., and Peter Van Inwagen. Earlier drafts were read at the University of Chicago, Wayne State, and as an Invited Paper at the American Philosophical Association Western Division meetings, 1976. Comments and criticisms by Ernest Sosa, the official respondent at the APA meeting, were exceptionally helpful.

BOOKS RECEIVED

- Evandro Agazzi, *Il Concetto Di Progresso Nella Scienze* (Milano: Feltrinelli Editore, 1976), 169 pp.
- Nicholas Capaldi and Luis E. Navia (eds.), *Journeys Through Philosophy* (Buffalo: Prometheus Books, 1977), 471 pp.
- D. P. Chattopadhyaya, *History, Society and Polity* (New Delhi: Macmillan Company of India Ltd., 1976), 281 pp.
- Lars O. Ericsson, *Justice in the Distribution of Economic Resources* (Stockholm: Almqvist & Wiksell International, 1976), 150 pp.
- Antony Flew, *Thinking Straight* (Buffalo: Prometheus Books, 1977), 127 pp.
- K. E. Goodpaster (ed.), *Perspectives on Morality* (Notre Dame: University of Notre Dame Press, 1976), 244 pp.
- Ted Honderich, *Political Violence* (Ithaca: Cornell University Press, 1977), 118 pp.
- Howard P. Kainz, *Hegel's Phenomenology, Part I: Analysis and Commentary* (University: The University of Alabama Press, 1976), 218 pp.
- W. Kang, *G. H. Mead's Concept of Rationality* (The Hague: Mouton Publishers, 1976), 219 pp.
- Roger A. Kenyon, *Existential Structures* (New York: Philosophical Library Inc., 1976), 63 pp.
- John King-Farlow and William R. Shea (eds.), *Values and the Quality of Life* (New York: Science History Publications, 1976), 181 pp.
- Reinhard Kleinknecht, *Ekkehard Wüst: Lehrbuch der elementaren Logik, Band 1* (München: Deutscher Taschenbuch Verlag, 1976), 256 pp.
- , *Ekkehard Wüst: Lehrbuch der elementaren Logik, Band 2* (München: Deutscher Taschenbuch Verlag, 1976), 230 pp.
- V. H. Klenk, *Wittgenstein's Philosophy of Mathematics* (The Hague: Martinus Nijhoff, 1976), 126 pp.
- Eberhard Knobloch, *Die Mathematischen Studien Von G. W. Leibniz zur Kombinatorik* (Wiesbaden: Franz Steiner Verlag GMBH, 1976), 339 pp. (*Studia Leibnitiana* Supplementary Volume XVI).
- A. Lamacchia et al., *Per Una Storia Della Critica Del Conoscere* (Bari: Ecumenica Editrice, 1976), 217 pp.
- Lars Lindahl, *Position and Change* (Dordrecht: D. Reidel, 1977), 299 pp.
- Norman Malcolm, *Memory and Mind* (Ithaca: Cornell University Press, 1977), 277 pp.
- Peter J. McCormick, *Heidegger and the Language of the World* (Ottawa: University of Ottawa Press, 1976), 208 pp.
- Charles Moraze, *The Logic of History* (The Hague: Mouton Publishers, 1976), 260 pp.
- Leonce Paquet, *Les Cyniques Grecs* (Ottawa: University of Ottawa Press, 1975), 304 pp.
- John L. Pollock, *Subjunctive Reasoning* (Dordrecht and Boston: D. Reidel, 1976), 255 pp.
- James Rachels (ed.), *Understanding Moral Philosophy* (Encino and Belmont: Dickenson Publishing Co., Inc., 1976), 417 pp.
- Nicholas Rescher, *Plausible Reasoning* (Assen: Royal Vangorcum Ltd., 1976), 124 pp.
- Amelie Oksenberg Rorty (ed.), *The Identities of Persons* (Berkeley: University of California Press, 1976), 333 pp.
- Albert Rothenberg and Carl R. Hausman (eds.), *The Creativity Question* (Durham: Duke University Press, 1976), 366 pp.
- Gilbert Ryle, *Contemporary Aspects of Philosophy* (Stocksfield: Oriel Press, 1977), 300 pp.
- Marshall Sahlins, *Culture and Practical Reason* (Chicago: University of Chicago Press, 1976), 252 pp.
- Oswald Schwemmer, *Theorie der rationalen Erklärung* (München: Verlag C. H. Beck, 1976), 276 pp.
- Huston Smith, *Forgotten Truth* (New York: Harper and Row, 1976), 182 pp.
- Torbjörn Tännsjö, *The Relevance of Metaethics to Ethics* (Stockholm: Almqvist & Wiksell International, 1976), 226 pp.
- Rainer W. Trapp, *Analytische Ontologie* (Frankfurt: Vittorio Klostermann, 1976), 261 pp.
- Klaus Wagn, *What Time Does* (München: Caann Verlag, 1976), 95 pp.

CORRIGENDUM

Volume 14 (1977)

JOHN COOPER: *The Psychology of Justice in Plato* (pp. 151-157).

Page 154, col. 1, line 38: insert "or kinds" between "classes" and "of individuals."

Volume LXXXV

MIND—OCTOBER 1977

Number 340

Edited by D. W. HAMLYN

Articles:

Institutions, Practices and Moral Rules
Hume on Moral Agreement
Wittgenstein and Materialism
Plato and Aristotle on Friendship and Altruism
Scepticism and Causal Theories of Knowledge

H. KHATCHADOURIAN
H. JENSEN
J. F. M. HUNTER
J. ANNAS
A. J. HOLLAND

Discussions:

Central Illocutionary Force and Meaning
Promises and Threats
Another Moral Standard
Goldman on the Non-contingency Thesis
Morality and Value
Two Forms of Theorizing About Law

J. R. BOATRIGET
V. PEETZ
T. W. LOCKHAET
R. A. SHINER
R. SITELMAN
J. G. HUND

Critical Notice:

Linguistic Behaviour by Jonathan Bennett

B. HARRISON

Book Reviews

Books Received

Notice

Published for THE MIND ASSOCIATION by BASIL BLACKWELL

Annual Subscription:

Institutions—£7.00 (inland and overseas), \$14.00 (USA and Canada)
Individuals—£4.00 (inland), £6.00 (overseas), \$12.00 (USA and Canada)
Single issue: £2.00 (inland and overseas), \$4.00 (USA and Canada)

Social Theory and Practice



An International and Interdisciplinary Journal of Social Philosophy

Volume 4, Number 2

Spring 1977

James M. Humber — Abortion, Fetal Research, and the Law

George Sher — An Unsolved Problem About Punishment

Robert F. Ladenson — Mill's Concept of Individuality

Fred R. Berger — Pornography, Sex, and Censorship

Hardy Jones — Fairness, Meritocracy, and Reverse Discrimination

Alison Jagger — Relaxing the Limits on Preferential Treatment

Social Theory and Practice is published twice yearly, fall and spring, by the Department of Philosophy, Florida State University. Four issues constitute a volume. Rates are \$10 per volume for individuals, \$21 for institutions. Manuscripts and subscription requests should be sent to *Social Theory and Practice*, Department of Philosophy, Florida State University, Tallahassee, FL 32306.

GRAZ PHILOSOPHISCHE STUDIEN

BAND 2

1976

VOLUME 2

Aufsätze

Articles

- Theorienstruktur in der Literaturwissenschaft Heide GÜTTNER
 Interpretation und Zielsetzung. Betrachtungen zum Problem
 der Eigenart der juristischen Interpretation Ota WEINBERGER
 Knowledge and Belief; Facts and Propositions Joseph MARGOLIS
 Knowledge, Belief, Existence, and Quantifiers — A Note on Hintikka Wolfgang LENZEN
 Über Wahrnehmung von Aspekten Petra von MORSTEIN
 The Factuality of Facts Reinhardt GROSSMANN
 On the Phenomenology of Partial Judgment John M. VICKERS
 From Purity to Reference Ronald SCALES
 Ad Baculum John WOODS & Douglas WALTON
 Das normative Fundament der Sprache: Ja und Aber Eike von SAVIGNY
 Handlungsgraphen. Graphentheoretische Modelle in der analytischen Handlungsphilosophie Hans LENK
 Laws in the Social Sciences V. Warren BOURGEOIS
 'Intentionale Erklärungen'. Kritische Gedanken zu Henrik von Wrights
 Sicht der Erklärung Reinhold BICHLER

Besprechungen

Review Articles

- Hubert SCHLEICHERT: *Logischer Empirismus — der Wiener Kreis*,
 München: Fink 1975 Reinhard KAMITZ
 Ernst TOPITSCH: *Die Voraussetzungen der Transzendentalphilosophie*.
 Kant in weltanschauungsanalytischer Beleuchtung, Hamburg:
 Hoffmann und Campe 1975 Stephan KÜRNER
 Edo PIVCEVIC (Hrsg.): *Phenomenology and Philosophical Understanding*,
 London: Cambridge University Press 1975 Lothar ELEY
 Alvin PLANTINGA: *The Nature of Necessity*, London:
 Oxford University Press 1974 John B. BACON
 Karl ACHAM: *Analytische Geschichtsphilosophie*. Eine kritische Einführung,
 Freiburg: Alber 1974 Herbert SCHNADELBACH

Herausgeber

Editor

Prof. Rudolf HALLER, Philosophisches Institut, Universität Graz, Heinrichstraße 26, A-8010, Graz, Österreich/Austria.

the philosophical forum A QUARTERLY

Volume VII, No. 2
Marx W. Wartofsky, Editor

- Nikhil Bhattacharya** John Dewey's Philosophy of Science
Barbara V. E. Klein Mind-Body Identity Relativized
Rebecca R. DeBoer Cartesian Categories in Mind-Body Identity Theories
Kenneth Dorter The Dialectic of Plato's Method of Hypothesis
Joel Peterson Problems in the Sartrean Paradigm of Life as a Project
William Sellers Discussion: Skepticism and Inquiry

Mail to: P.O. Box 247
 Boston University Station
 Boston, Massachusetts 02215 U.S.A.

the review of metaphysics

a philosophical quarterly

SEPTEMBER 1977 | VOL. XXXI, No. 1 | ISSUE No. 121 | \$5.00

articles

ROBERT SOKOLOWSKI

RICHARD J. BERNSTEIN

MERRIE BERGMANN

JOSEPH L. BLAU

Picturing

Why Hegel Now?

Logic and Sortal Incorrectness

Emerson's Transcendentalist Individualism as a
Social Philosophy

critical study

MAURICE A. FINOCCHIARO

The Uses of History in the Interpretation of Science

discussion

M. J. INWOOD

Solomon, Hegel, and Truth

philosophical abstracts

doctoral dissertations

visiting philosophers from abroad

north american philosophers abroad

professors emeriti

announcements

Individual Subscriptions \$12.00

Institutional Subscription \$18.00

Student Subscriptions \$3.00

Jude P. Dougherty, Editor

The Catholic University of America, Washington, D. C. 20064
Twenty-Year Cumulative Index, 1947-1967 is available for \$3.00 a copy.

THE PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 27, No. 109, OCTOBER 1977

ARTICLES

PLATO ON NAMING

GAIL FINE

AGAINST AN ORTHODOX

INTERPRETATION OF HOBBS W. E. LYONS

KRIPKE ON IDENTITY AND NECESSITY

DAVID BOSTOCK

'WHY DO THINK THAT ...?' QUESTIONS

EUGENE VALBERG

DISCUSSION

FREY ON INTERESTS AND ANIMAL RIGHTS

TOM REGAN

CRITICAL STUDY

JONATHAN BENNETT:

LINGUISTIC BEHAVIOUR

ROEIM CAMERON

BOOK REVIEWS

BOOKS RECEIVED

*Edited in the University of St Andrews by Bernard Mayo with the assistance of Christopher Bryant, Anthony Ellis
and Stephen Read*

Published for the Scots Philosophical Club by the University of St Andrews

Annual Subscriptions:

Individuals £4.00 (\$10.50)

Institutions £7.00 (\$17.00)

Single Issues £1.25 or \$3.00

(Institutions £2.00 or \$5.00)

for orders placed directly with the

Journals Manager, Scottish Academic Press, Ltd., 33 Montgomery St., Edinburgh EH7 5JX, Scotland

The British Journal for the Philosophy of Science

Contents, Volume 28 Number 1 March 1977

Articles

- JOHN MAYBERRY On the Consistency Problem for Set Theory: An Essay on the Cantorian Foundations of Classical Mathematics (I)
ALFONS GRIEDER Relativity, Causality and the 'Substratum'
FERREL CHRISTENSEN How to Establish Non-Conventional Isochrony
JON DORLING The Eliminability of Masses and Forces in Newtonian Particle Mechanics: Suppes Reconsidered

Discussion Note

- PETER URBACH Putnam and Wilks and Mind and Body

Review Articles

- MICHAEL REDHEAD Wave-Particle Duality (Review of M. Audi: *The Interpretation of Quantum Mechanics* and J. M. Jauch: *Are Quanta Real?*)
FRED D'AGOSTINO Knowledge of Language (Review of D. Cooper *Knowledge of Language*)

Reviews

- W. V. QUINE on I. Lakatos: *Proofs and Refutations: The Logic of Mathematical Discovery*;
SUSAN HAACK on D. J. O'Connor: *The Correspondence Theory of Truth* and C. J. F. Williams: *What is Truth?*; TOM SETTLE on M. Bunge: *Method, Model and Matter*;
JOHN BELL on R. C. Jeffrey and G. S. Boolos: *Computability and Logic*

Announcements

**Published for the British Society for the Philosophy of Science by
Aberdeen University Press, Farmers Hall, Aberdeen AB9 2XT, Scotland**

Editors: John Watkins and John Worrall

Price £2.50 (\$8.00)

Subscription Price £7.00 (\$22.00)

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

in cooperation with a distinguished group of American
and foreign scholars

Descriptive, analytic, critical, and historical articles representing the major contemporary trends in philosophy. In addition to phenomenological research, PPR publishes studies in a wide range of areas including ethics and value theory, metaphysics, aesthetics, logic, language, political, social, and religious philosophy, theory of knowledge, and the sciences.

Address all communications to
Philosophy and Phenomenological Research
State University of New York at Buffalo
320 Baldy Hall
Buffalo, New York 14260

Annual subscription rate, \$11.00 for libraries and institutions, \$9.00 for individuals; single copies \$2.75 and \$2.25 respectively, plus postage and handling on back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed, stamped envelope, or equivalent postage.

A circular listing the main contents of the Journal since 1940 will be sent upon request.

The Monist *An International Quarterly Journal of General Philosophical Inquiry*

Founded 1888 by **Edward C. Hegeler**

Editor, Eugene FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal, William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor, Ann FREEMAN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate with the Editor in advance for Special Instructions. Papers should be 4000-8000 words in length and must be submitted in duplicate nine months prior to the scheduled publication of the issue, accompanied by return postage.

SCHEDULED PUBLICATION DATES:

Vol. 61,	No. 1	Jan.,	1978	Conceptions of the Self: East and West
Vol. 61,	No. 2	Apr.,	1978	The Philosophy of Thomas Reid
Vol. 61,	No. 3	July,	1978	Nominalism, Past and Present
Vol. 61,	No. 4	Oct.,	1978	The Philosophy and Psychology of Cognition
Vol. 62,	No. 1	Jan.,	1979	Parmenides and His Influence on Western Thought
Vol. 62,	No. 2	Apr.,	1979	Truth, Meaning, and Reference
Vol. 62,	No. 3	July,	1979	The Concept of A Person in Ethical Theory
Vol. 62,	No. 4	Oct.,	1979	Objectivity in Knowledge and Valuation

Editorial Office: Box 1908, Los Gatos, California 95030

Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, \$18.00; Individuals, \$12.00. Single copies: \$4.50. Foreign postage: Add 25 cents to single copy rate or \$1.00 to subscription rate.

INQUIRY

Editor: Alastair Hannay

Founding and Consulting Editor: Arne Naess

Volume 20 (1977), No. 1 (Spring)

Marković's Concept of *Praxis* as Norm
Punishing the Innocent — Unintentionally
The Limited Neutrality of Typologies of Systems: A Reply to Gullvåg
The Rationality Principle and Action Explanations:
Koertge's Reconstruction of Popper's Logic of Action Explanations

David A. Crocke-
Alan Wertheimer:
Arne Naess

Peter Glück
and Michael Schmid

The Liberation of Nature? (Christopher Stone: *Should Trees Have Standing?*; Peter Singer: *Animal Liberation*)
Attitudes and Action (Stephan Körner: *Experience and Conduct*)

John Rodman
R. A. Sharpe

Volume 20 (1977), No. 2 (Summer)

Description and Expression: Physicalism Restricted
Berkeley, Truth, and the World
Frege's Alleged Realism
Feyerabend's Discourse against Method: A Marxist Critique

Virgil C. Aldrich
Eric Bush
Hans D. Sluga
J. Curthoys
and W. Suchting

Subscription: Institutions — N.kr. 90.00, £6.25, \$18.00, postage included.
Individuals — N.kr. 60.00, £5.00, \$12.00.

Published quarterly by UNIVERSITETSFORLAGET, P.O. Box 7508, Skillebeek, Oslo 2, Norway — P.O. Box 142, Boston, Mass. 02113, USA

AMERICAN PHILOSOPHICAL QUARTERLY

PRIZE ESSAY FOR 1978

Topic:

THE ETHICS OF BELIEF:
WHY SHOULD I BE RATIONAL?

Deadline:

JANUARY 15, 1978

Prize:

\$100 and publication in the OCTOBER 1978 issue of the APQ

Conditions:

Only current subscribers to the APQ are eligible to make submissions. These must be written in English. They should be presented in the standard APQ typescript format and must not exceed 7,000 words in length. Entries will be judged by a panel of readers appointed by the editor of the APQ. The right of declining to make any award is reserved if no submissions of suitable quality are received.